

## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur : ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite de ce travail expose à des poursuites pénales.

Contact : [portail-publi@ut-capitole.fr](mailto:portail-publi@ut-capitole.fr)

## LIENS

Code la Propriété Intellectuelle – Articles L. 122-4 et L. 335-1 à L. 335-10

Loi n° 92-597 du 1<sup>er</sup> juillet 1992, publiée au *Journal Officiel* du 2 juillet 1992

<http://www.cfcopies.com/V2/leg/leg-droi.php>

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



# THÈSE



En vue de l'obtention du

**DOCTORAT DE L'UNIVERSITE DE TOULOUSE**

**Délivré par l'Université Toulouse Capitole**

École doctorale : **Sciences Economiques-Toulouse School of Economics**

---

**Présentée et soutenue par**

**LAPENTA Elia**

**le 28 September 2020**

**Three Essays in Hypothesis Testing**

---

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R (UMR CNRS 5314 – INRA 1415)**

**Directeur de thèse** : Mons. Pascal Lavergne, Professeur,  
École d'Économie de Toulouse, Université Toulouse Capitole

**JURY**

**Rapporteurs** M.me, Ingrid Van Keilegom, Professeur, KU University, Leuven;  
Mons., Juan-Carlos, Escanciano, Professeur, University Carlos III, Madrid;

**Suffragants** Mons. Jean-Pierre Florens, Professeur, École d'Économie de Toulouse,  
Université Toulouse Capitole;



# Introduction

This thesis contains three chapters in Hypothesis Testing for semi and non parametric models. The common features of these chapters are two. First, testing is based on the bootstrap. The test statistics proposed are not asymptotically pivotal. Their null asymptotic distributions are difficult to compute, so they cannot be used for the computation of the critical values. The bootstrap, instead, allows obtaining the critical values in a relatively simple way. All bootstrap tests constructed in this work exploit the information under the null hypotheses.

The second common feature across the works in this thesis is the employment of bias corrections for the computation of the statistics. The testing frameworks are non or semi parametric, so the estimators employed are biased. In these contexts, the use of bias corrections allows improving the performance of the tests. Intuitively, the need to control for estimation bias requires shrinking the set of tuning parameters (bandwidths) admissible for inference. The introduction of bias corrections alleviates this problem, enlarging the set of tuning parameters admissible for testing. This makes the tests more robust to the choice of such parameters. It also allows inference using selection rules that are not admissible without bias corrections, avoiding undersmoothing.

The First and the Third chapter develop new tests for models containing nonparametrically generated variables. Such variables are not observed by the researcher but are nonparametrically identified and estimable. In the Second chapter, testing is developed in a framework where all regressors are observed, but inference employs an iterative bias correction method known as  $L_2$  boosting. This method extends the bias correction used in the first chapter, and its employment in testing is novel in the literature.

The contributions of this thesis are threefold. First, new tests are developed for models involving generated variables. The econometric/statistical literature has mainly focused on estimation for these models, but testing appears to be a relatively unexplored area. Second, the works provide new bootstrap procedures for models involving generated regressors. These procedures are new in the literature, as they need to mimic the estimation error coming from estimating the unobserved variables. Third, bias corrections are implemented for inference in models with and without generated regressors. The use of such bias corrections is novel for the testing problems considered.

The First chapter (*A Bootstrap Specification Test for Semiparametric Models with Generated Regressors*) provides a specification test for semiparametric models with nonparametrically generated regressors. Applications include models with endogenous regressors identified by control functions, semiparametric sample-selection models, or binary games with incomplete information. The statistic proposed is a Cramer Von Mises statistic built from the residuals of the semiparametric model. Due to the presence of generated regressors, the bootstrap tests available in the literature cannot be employed in this context. In particular, the presence of the generated variables implies extra terms in the asymptotic expansion of the statistic, as such variables need to be estimated. Since the bootstrap procedures available in the literature are developed for models where all variables are observed, they cannot mimic such extra terms.

The main contribution of this chapter is to develop a novel wild-bootstrap procedure and to show its validity. The test involves bias corrections of the nonparametric estimators, thus undersmoothing is avoided. To the best of my knowledge, the employment of bias corrections is novel for inference in models involving generated variables.

The Second chapter (*A Nonparametric Encompassing Test*) is co-authored with Prof. Pascal Lavergne. It develops a test based on the encompassing principle to choose between two alternative models. According to the encompassing principle, a model  $\mathcal{M}_1$  encompasses a model  $\mathcal{M}_2$  if  $\mathcal{M}_1$  can explain the results of  $\mathcal{M}_2$ . This is a natural principle used in science to choose between two alternative theories: a new theory can replace an older one if it explains both new phenomena and the old phenomena explained by the older theory. The encompassing tests in the current literature either rely on parametric functional forms or, when relying on nonparametric specifications, they condition the analysis on fixed values of the explanatory variables. This chapter provides a nonparametric encompassing test. The procedure developed does not rely on neither functional forms nor on specific values of the explanatory variables. The statistic is computed according to the  $L_2$  boosting algorithm which allows to obtain a good robustness of the test with respect to the choice of the smoothing parameter. The critical values are simulated by a wild-bootstrap procedure which is proven to be valid in the presence of  $L_2$  boosting iterations.

Finally, the Third chapter (*Testing Bayesian-Nash Behavior in Binary Games with Incomplete Information and Correlated Types*), also co-authored with Prof. Pascal Lavergne, develops a test for checking if the distribution of the observed data can be characterized by a unique Bayesian-Nash equilibrium. The framework is a binary game with incomplete information, where agents' types are allowed to be mutually correlated. The usefulness of this test is based on two points. First, the uniqueness of the Bayesian-Nash equilibrium is key to identify the fundamentals of the game. Second, testing for a Bayesian-Nash behavior is interesting *per se*, as it is an assumption often postulated in game-theoretical models. The test proposed relies on rationalization results in Liu et al. (2017). From an econometric point of view, the model involves generated regressors, the conditional probabilities concerning agents' strategies. The test statistic is constructed by extending the  $L_2$  boosting procedure used in the second chapter to a context with generated variables. This boosting procedure is effective to control the estimation bias arising in this context. Since the asymptotic distribution of the statistic depends on unknown features of the data, a novel Multinomial Bootstrap procedure is constructed to obtain the critical value, and its validity is proved. This procedure resamples the observations imposing that a unique Bayesian-Nash equilibrium is played and preserves the binary nature of the agents' decisions in the "bootstrap world". The Multinomial Bootstrap procedure developed appears to be new in the literature for two reasons. First, multinomial bootstrap schemes resampling the data under the null hypothesis have not been explored in the statistical/econometric literature. Second, the validity of the multinomial bootstrap is obtained in the presence of generated variables.

**Acknowledgement** I am highly indebted to Prof. Pascal Lavergne who has provided a unique and invaluable guidance to the work behind this thesis. I really thank him for all his patience and support as well as for the precious insights he has given to me during my PhD. A special thank goes to Prof. Jean-Pierre Florens who has given to me great many insightful comments and suggestions, and from whom I have benefited of many precious discussions. I also thank Prof. Ingrid Van Keilegom for her comments and suggestions given to me during her visits in Toulouse, and Prof. Juan Carlos Escanciano for his precious discussions and insights I have benefited from during my visit at the University Carlos III.

The Econometrics Research Group at TSE has provided a greatly stimulating environment for conducting the work in this thesis, and I am really thankful to all the components of the group. In particular, I would like to thank Anne Vanhem, Thierry Magnac, Andrii Babii, Nour Meddahi, Eric Gautier, Abde-laati Dahouia, Cristina Gualdani, Christian Bontemp, Olivier de Groote, Francois Poinas, and Christophe Gaillac.

Finally, my experience in Toulouse and TSE would not have been so enjoyable without the great friends I have had during the PhD years. They have given to me deep moral support when I was getting stuck with my research, as well as funny and cool moments. In particular, I would like to thank, among others, Paloma Carrillo, Matheus Bueno, Filippo d’Arcangelo, Jacint Enrich, Alberto Grillo, Miguel Zere-cero, Miren Azkatarate, and Marica Valente. Finally, I would like to thank my parents, Bruno and Caterina, and my brothers, Gaetano and Federico, for all their support, despite the distance

# Résumé

Cette thèse se compose de trois chapitres portant sur les tests des hypothèses pour des modèles semi- et non-paramétriques. Ces chapitres ont deux caractéristiques en commun. Premièrement, les tests sont basés sur le *bootstrap*. Les statistiques présentées ne sont pas asymptotiquement pivotales et leurs distributions asymptotiques dépendent de paramètres inconnus. De surcroît, la computation des valeurs critiques basée sur ces distributions est généralement difficile. Pour cela, l'approximation asymptotique ne constitue pas une voie attractive. Le *bootstrap*, cependant, permet d'obtenir les valeurs critiques de façon relativement simple. La deuxième caractéristique liant les travaux de cette thèse réside dans l'utilisation de corrections de biais pour la computation des statistiques. Étant donné que les contextes analysés sont non- ou semi- paramétriques, les estimateurs utilisés sont biaisés. Dans ces cas, l'utilisation de telles corrections permet d'améliorer les performances des tests. Intuitivement, le besoin de contrôler les biais d'estimation impose des restrictions relativement fortes sur les *bandwidths* qui peuvent être employées pour la computation des estimateurs. L'introduction des corrections de biais assouplit ces problèmes, élargissant l'ensemble des *bandwidths* pouvant être utilisées pour le test. Cette caractéristique renforce le test par rapport au choix de ces paramètres et permet d'employer des règles de sélection pour l'inférence qui n'étaient pas admissibles sans lesdites corrections, évitant le phénomène de *undersmoothing*.

Le premier et le troisième chapitres développent de nouveaux tests pour des modèles contenant variables non-paramétriquement générées. De telles variables ne sont pas observées par les chercheurs mais sont cependant identifiées et peuvent être estimées. Dans le deuxième chapitre, nous développons des tests dans un contexte où toutes les variables sont observées mais l'inférence utilise une correction de biais itérative connue comme *L<sub>2</sub>-boosting*. Cette méthode étend la correction de biais employée dans le premier chapitre et son utilisation dans un test d'hypothèse non-paramétrique est nouveau dans la littérature.

Cette thèse apporte principalement trois contributions. Premièrement, cette étude développe de nouveaux tests pour des modèles comprenant variables générées. La littérature statistique/économétrique s'est focalisée sur l'estimation de ces modèles, mais le problème des tests est relativement moins exploré. Deuxièmement, elle propose de nouvelles procédures de *bootstrap* pour des modèles comprenant régresseurs générés. Ces procédures sont nouvelles dans la littérature car elles doivent répliquer l'erreur d'estimation provenant des régresseurs générés. Enfin, des corrections de biais sont employées pour l'inférence dans des modèles avec et sans régresseurs générés. L'utilisation de telles corrections de biais constitue une nouveauté dans les problèmes d'inférence considérés.

Le premier chapitre (*A Bootstrap Specification Test for Semiparametric Models with Generated Regressors*) fournit un test de spécification pour des modèles semi-paramétriques contenant variables non-paramétriquement générées. Les applications de ce test inclut des modèles avec régresseurs endogènes où les composantes structurelles sont identifiées par des fonctions de contrôle, ainsi que des modèles semi-paramétriques avec sélection dans l'échantillon et des modèles de jeux binaires avec information

incomplète. Les statistiques proposées sont du type Cramer-Von Mises. Celles-ci sont obtenues des résidus du modèle semi-paramétrique. À cause des régresseurs générés, les tests basés sur le bootstrap disponibles dans la littérature ne peuvent pas être employés dans ce contexte. En particulier, la présence des variables générées implique des termes additionnels dans l'expansion asymptotique de la statistique, car de telles variables doivent être estimées. Étant donné que les procédures de rééchantillonnage disponibles dans la littérature sont développées pour modèles où toutes les variables sont observées, elles ne peuvent pas répliquer les fluctuations de tels termes additionnels. La contribution principale de ce chapitre est de développer une nouvelle procédure de rééchantillonnage et prouver sa validité. L'implémentation du test emploie des correction de biais et le phénomène de *undersmoothing* est par conséquent évité.

Le deuxième chapitre (*A Nonparametric Encompassing Test*) a été développé en collaboration avec Prof. Pascal Lavergne. Ce chapitre propose un test basé sur le principe de l'*Encompassing* permettant de choisir entre deux modèles alternatifs. Suivant le principe de l'*Encompassing*, un modèle  $\mathcal{M}_1$  comprend un modèle  $\mathcal{M}_2$  si  $\mathcal{M}_1$  peut expliquer les résultats de  $\mathcal{M}_2$ . Celui-ci est un principe naturel utilisé en science pour choisir entre deux théories alternatives : une nouvelle théorie peut en remplacer une autre si la première explique les "nouveaux" aussi bien que les "vieux" phénomènes déjà expliqués par la vieille théorie. En l'état actuel, les tests d'*Encompassing* présents dans la littérature sont basés sur des formes fonctionnelles paramétriques ou, lorsqu'ils prennent en compte des modèles non-paramétriques, considèrent des valeurs fixes des variables explicatives. Ce deuxième chapitre développe un test d'*Encompassing* non-paramétrique dans lequel la procédure proposée n'est pas basée sur des valeurs fixes des variables explicatives ou sur des formes fonctionnelles spécifiques. La statistique est calculée en utilisant l'algorithme de *L<sub>2</sub>-boosting* qui permet d'obtenir une robustesse satisfaisante du test par rapport au choix de la bandwidth. Les valeurs critiques sont calculées avec une procédure de *Wild Bootstrap* la validité de laquelle est démontrée avec un nombre arbitraire d'itérations de *L<sub>2</sub>-boosting*.

Pour conclure, le troisième chapitre (*Testing Bayesian-Nash Behavior in Binary Games with Incomplete Information and Correlated Types*), aussi développé avec Prof. Pascal Lavergne, propose un test pour vérifier statistiquement si la distribution conditionnelle des données observées est caractérisée par un unique équilibre Nash-Bayésien. Le contexte est un jeu binaire avec information incomplète où les typologies d'agents sont mutuellement corrélées. L'utilité de ce test est basée sur deux points. Premièrement, l'unicité de l'équilibre Nash-Bayésien est crucial pour obtenir l'identification des fondamentaux du modèle. Deuxièmement, la vérification du comportement Nash-Bayésien est intéressant *per se*, car il s'agit d'une hypothèse souvent postulée dans les modèles de jeux microéconomiques. Le test proposé est basé sur des résultats de rationalisations fournis par Liu et al. (2017). D'une perspective purement économétrique, le contexte comprend des variables non-paramétriquement générées correspondant aux probabilités conditionnelles des stratégies des agents. La statistique est obtenue en appliquant le principe de *L<sub>2</sub>-boosting* utilisé dans le deuxième chapitre dans un contexte avec variables générées. Cette procédure de boosting est efficace pour contrôler le biais provenant de l'estimation des



probabilités conditionnelles. Puisque la distribution asymptotique dépend par des caractéristiques inconnues des données, une nouvelle procédure de bootstrap multinomial est développée afin d'obtenir les valeurs critiques du test et sa validité est démontrée. La procédure impose la présence d'un unique équilibre Nash-Bayésien dans le rééchantillonnage et permet de préserver la nature binaire des choix des agents dans le "bootstrap world". Une telle procédure de rééchantillonnage représente une nouveauté pour deux raisons. Premièrement, les procédures de rééchantillonnage multinomial développées dans la littérature n'imposent pas l'hypothèse testée. Deuxièmement, la validité de cette procédure est obtenue en présence de variables générées.

**Remerciements.** Je suis énormément reconnaissant envers Prof. Pascal Lavergne qui a été un guide unique et exceptionnel pour les travaux derrière cette thèse. Je le remercie sincèrement pour toute sa patience ainsi que pour le soutien qu'il m'a donné et pour les précieuses directions tout au long de mon doctorat. Un remerciement spécial va aussi à Prof. Jean-Pierre Florens qui a fourni de nombreux commentaires, suggestions et direction desquelles j'ai énormément bénéficié. Je voudrais aussi sincèrement remercier Prof. Ingrid Van Keilegom pour ses commentaires et suggestions pendant ses visites à Toulouse et Prof. Juan Carlos Escanciano pour les précieuses discussions pendant ma visite à l'Université Carlos III de Madrid. Le Groupe de Recherche en Économétrie de l'École d'Économie de Toulouse a fourni un cadre de travail exceptionnel pour développer cette thèse. Je suis très reconnaissant envers tous les membres du groupe. En particulier, je voudrais remercier Anne Vanhem, Thierry Magnac, Andrii Babii, Nour Meddahi, Eric Gautier, Abdelaati Dahouia, Cristina Gualdani, Christian Bontemps, Olivier de Groote, François Poinas et Christophe Gaillac. Mon expérience à Toulouse n'aurait été aussi pleine et enrichissante sans les amis exceptionnels que j'ai eu le plaisir de rencontrer pendant mon doctorat. Je voudrais remercier Paloma Carrillo, Matheus Bueno, Filippo d'Arcangelo, Jacint Enrich, Alberto Grillo, Miguel Zerecero, Miren Azkatarate et Marica Valente. Finalement, je voudrais remercier mes parents, Bruno et Caterina, et mes frères, Gaetano et Federico, pour tout leur soutien malgré la distance.



# Contents

|          |   |            |
|----------|---|------------|
| <b>1</b> | <b>A Bootstrap Specification Test for Semiparametric Models with Generated Regressors</b> | <b>1</b>   |
| <b>2</b> | <b>A Nonparametric Encompassing Test</b>  | <b>71</b>  |
| <b>3</b> | <b>Testing Bayesian Nash Behavior in Binary Games with Incomplete Information and</b>     |            |
|          | <b>Correlated Types</b>   | <b>102</b> |



# Chapter 1: A Bootstrap Specification Test for Semiparametric Models with Generated Regressors

Elia Lapenta \*

Toulouse School of Economics

Job Market Paper

## Abstract

This paper provides a specification test for semiparametric models with nonparametrically generated regressors. Such variables are not observed by the researcher but are nonparametrically identified and estimable. Applications of the test include models with endogenous regressors identified by control functions, semiparametric sample selection models, or binary games with incomplete information. The statistic is built from the residuals of the semiparametric model, and a novel wild bootstrap procedure is shown to provide valid critical values. We consider nonparametric estimators with an automatic bias correction that makes the test implementable without undersmoothing. In simulations the test exhibits good small-sample performances, and an application to women's labor force participation decisions shows the implementation of the test in a real-data context.

**Keywords :** Hypothesis Testing, Bootstrap, Semiparametric Model, Control Function.

**JEL Classification :** C01, C12, C14

---

\*I am very grateful to Pascal Lavergne for many valuable insights on the work behind this paper. I am also very grateful to Jean-Pierre Florens, Juan Carlos Escanciano, and Ingrid Van Keilegom for many useful insights, suggestions, and comments. I also thank Andrii Babii, Christian Bontemps, Matheus de Souza Bueno, Paloma Carrillo, Eric Gautier, Cristina Gualdani, Emmanuel Guerre, Jihyun Kim, Louise Laage, Thierry Magnac, Enno Mammen, Nour Meddahi, Christophe Rothe, Melanie Schienle, Sami Stouli, Lorenzo Trapani, Marica Valente, and Anne Vanhems for their comments and suggestions. This research has received financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement N°295298.

# 1 Introduction

A strong tradition on specification testing is present in econometrics. In this paper, we contribute to the literature by providing a specification test for semiparametric models with nonparametrically generated variables. These variables are not observed by the researcher but are nonparametrically identified and estimable. Examples of semiparametric models with nonparametrically generated variables are common in empirical frameworks and include endogenous models with control functions ([Blundell & Powell, 2004](#); [Newey \*et al.\*, 1999](#)), semiparametric sample-selection models ([Escanciano \*et al.\*, 2016](#)), or extensions of Tobit models ([Escanciano \*et al.\*, 2016](#)).

Checking the correct specification of a model is empirically relevant, as a misspecified model yields biased and inconsistent estimates and provides a misleading counterfactual analysis. If the models did not contain nonparametrically generated regressors, specification testing could be based on procedures already available in the econometric literature which assume that all the variables are observed (see e.g. [Delgado & Manteiga, 2001](#); [Xia \*et al.\*, 2004](#); [Fan & Li, 1996](#)). In principle, these tests could be naively applied by replacing the nonparametrically generated regressors with their estimates. Such a procedure, however, would deliver a wrong inference, as the nonparametrically generated variables must be estimated in a preliminary step, and this introduces an estimation error that needs to be taken into account.

The test we propose allows the researcher to check the validity of semiparametric moment conditions involving nonparametrically generated regressors. The statistic is based on a weighted sum of the residuals obtained from the estimation of the model and has a simple closed-form expression. We prove that such a statistic converges to a non-pivotal distribution, i.e. a distribution that depends on unknown features of the data generating process. Thus, to obtain the critical values necessary for testing, we develop a novel Wild Bootstrap scheme and prove its validity under low-level assumptions.

The wide range of application of this test comes from the widespread presence of generated regressors in empirical economic models. A first type of relevant setups includes semiparametric binary-choice models with endogeneity and control functions (see e.g. [Rivers & Vuong, 1988](#); [Newey \*et al.\*, 1999](#); [Blundell & Powell, 2004](#); [Imbens & Newey, 2009](#); [Blundell & Matzkin, 2014](#); [Wooldridge, 2015](#)). In these setups, the parameters of interest are generally not identified by standard IV assumptions, and the introduction of a control function allows to obtain identification (see [Blundell & Powell, 2004](#)). A further application of the test is provided by nonlinear semiparametric regressions with endogeneity that are separable with respect to an unobserved error term. [Newey \*et al.\* \(1999\)](#) highlights that the control function approach is convenient in such frameworks because it avoids an ill-posed inverse problem and allows a simple estimation procedure. Other examples of applications are semiparametric binary models with censoring and truncated models with selection, like double-hurdle models ([Escanciano \*et al.\*, 2016](#)). The test can also be applied to check the correct specification of semiparametric games with incomplete information, where agents make binary decisions, see [Aradillas-Lopez \*et al.\* \(2007\)](#), [Aradillas-Lopez \(2010\)](#), [Aradillas-Lopez \(2012\)](#), and [Lewbel & Tang \(2015\)](#).

The main contributions of this paper are threefold. First, we develop a specification test for the mentioned semiparametric models with nonparametrically generated variables. Second, we contribute to the literature on bootstrap inference by extending the Wild Bootstrap procedure ([Davidson & MacKinnon, 2010](#); [Delgado & Manteiga, 2001](#)) to a semiparametric context with nonparametrically generated regressors. Third, in the construction of the specification test, we adopt automatic bias corrections for the nonparametric estimators involved.

To build a specification test for semiparametric models with generated regressors, we start from the restriction they impose on the distribution of the data. In the benchmark example, such a restriction writes as a moment condition of the type

$$\begin{aligned} E\{Y | X, Z\} &= G_0(X^T \beta_0, V), \text{ where} \\ V &= X^e - m_0(Z), \end{aligned}$$

$(Y, X^e, X^T, Z^T)$  are observed,  $\beta_0$  is an unknown vector,  $G_0$  is a nonparametric function, and  $m_0(Z) = E\{X^e | Z\}$ .  $m_0$  is an unknown function. It is not restricted to have a specific functional form, so  $V$  is a nonparametrically generated regressor. To construct a test, we use an approach based on [Bierens & Ploberger \(1997\)](#) and [Stinchcombe & White \(1998\)](#). The test statistic proposed is a transformation of a weighted sum of the estimated residuals and has a simple closed-form expression. To provide tractable proofs, we obtain the residuals by kernel estimation.

For the nonparametric estimation of  $G_0$  and  $V$ , required to compute the residuals, we adopt an automatic bias correction. To the best of our knowledge, this is a novel approach in semiparametric models with generated regressors. This allows to widen the set of bandwidths admissible for the test compared to the case where such a bias correction is not used. Furthermore, it allows to implement the test by kernels of relatively low order. The bias correction employed is similar to the one used in [Xia \*et al.\* \(2004\)](#) who proposes a specification test for a single-index model where all the variables are observed. It can be considered a bootstrap estimate of the bias.

We show that asymptotically the statistic converges to a transformation of a Gaussian process. Since the asymptotic distribution is not pivotal, asymptotic tests are difficult to implement. Due to the presence of nonparametrically generated regressors, the asymptotic distribution involves nonparametric derivatives, so to obtain the asymptotic critical values one should estimate these nonparametric derivatives. This approach would suffer from two issues. First, the estimators of the nonparametric derivatives have a slow convergence rates, and this would seriously compromise the capacity of the test to provide a reliable inference in finite samples. Second, to estimate such derivatives a bandwidth should be selected, and this would introduce an additional parameter the researcher should set.

Instead, to compute the critical values we develop a novel two-step wild bootstrap procedure. In particular, we contribute to the literature on bootstrap inference by extending to a semiparametric context with nonparametrically generated variables the wild bootstrap scheme proposed in [Davidson & MacKinnon \(2010\)](#) for the linear two-stages least squares regression. We show the validity of this

wild bootstrap procedure under low-level conditions. The proofs for this result are challenging for two reasons. First, we need to handle a nonparametric estimator nesting inside another nonparametric estimator. Second, in the bootstrap context the estimators are computed with the “artificial data” resampled from the wild bootstrap scheme that contains estimation noise. As we show in a Monte Carlo simulation study, this test is able to provide a reliable inference in small samples, both in terms of size and power.

**Related literature.** A large literature analyzing the problem of estimation with nonparametrically generated regressors exists, beyond the studies already cited above. Early work includes [Pagan \(1984\)](#), [Ahn & Manski \(1993\)](#), [Ahn \(1997\)](#). More recent results are presented by [Li & Wooldridge \(2002\)](#), [Chen \*et al.\* \(2003\)](#), [Rothe \(2009\)](#), [Mammen \*et al.\* \(2012\)](#), [Mammen \*et al.\* \(2016\)](#), [Bravo \*et al.\* \(2018\)](#), [Hahn \*et al.\* \(2018\)](#), [Vanhems & Keilegom \(2019\)](#). The impact of generated regressors on the asymptotic distribution of a finite-dimensional estimator is analyzed in [Newey \(1994\)](#) and [Hahn & Ridder \(2013\)](#). [Escanciano \*et al.\* \(2014\)](#) obtains an expansion of the residuals from a regression involving some variables estimated in a preliminary step. These papers focus on estimation, and do not address specification testing in the presence of generated regressors.

From a methodological point of view, our work is closely related to [Escanciano \*et al.\* \(2014\)](#), with two main differences. First, in this paper we consider a bootstrap environment, while [Escanciano \*et al.\* \(2014\)](#) are not concerned with proving the validity of a bootstrap test. When proving the validity of the wild-bootstrap method, we need to handle the (uniform) convergence of a kernel estimator, where the projected variable is not observed but estimated. Thus, the results in [Escanciano \*et al.\* \(2014\)](#) are not applicable to the context of the present paper. Second, the sum of the residuals at the basis of our statistic contains bias-correction terms which are not present in their context. This allows to impose conditions on the bandwidths different from those they require and to avoid undersmoothing.

This study is also related to work on specification testing for semiparametric and nonparametric models, see [Fan & Li \(1996\)](#), [Chen \*et al.\* \(2003\)](#), [Ai & Chen \(2003\)](#), [Delgado \*et al.\* \(2006\)](#), [Einmahl & Van Keilegom \(2008\)](#), [Lavergne & Patilea \(2008\)](#), [Einmahl & Van Keilegom \(2008\)](#), [Delgado & Stute \(2008\)](#), [Escanciano & Song \(2010\)](#), [Lavergne \*et al.\* \(2015a\)](#), [Neumeyer & Van Keilegom \(2010\)](#). We use an approach similar to [Delgado & Manteiga \(2001\)](#), [Bierens & Ploberger \(1997\)](#), [Stinchcombe & White \(1998\)](#), only to cite a few. The difference with this literature comes from the presence of nonparametrically generated regressors which introduce extra terms in the asymptotic expansion.

**Organization of the paper.** Section 2 starts from a benchmark framework and describes the test in an intuitive way without delving into technical aspects. Section 3 defines in detail the estimation procedure employed to compute the residuals. Section 4 introduces the assumptions and obtains the asymptotic behavior of the statistic, while Section 5 sets up the bootstrap test and shows its validity. Section 6 contains other applications of the test, beyond the benchmark framework. Section 7 provides the simulation study assessing the small-sample behavior of the proposed test. It also contains an



empirical application to women’s labor force participation. Finally, Section [8](#) concludes. The Appendix contains the proofs for a general model encompassing all the examples considered.

## 2 The Test: Benchmark Example and Heuristics

We describe how to apply the test to a benchmark example. The presentation is kept at an intuitive level, without focusing on technicalities.

### 1. Binary-choice models with control functions

Let  $Y \in \{0, 1\}$  denote the discrete choice of an agent, and assume that such a choice depends on both a vector of covariates  $X \in \mathbb{R}^{dim(X)}$  and an unobserved error term  $u$  according to the model

$$Y = 1\{X^T \beta_0 \geq u\}, \quad (1)$$

where  $\beta_0 \in \mathbb{R}^{dim(X)}$  is an unknown vector and  $X = (X^e, Z_1)^T$ .  $X^e$  denotes the endogenous regressor correlated with  $u$ , while  $Z_1$  stands for the exogenous regressor. For simplicity, we assume that each of them is a scalar random variable, as the generalization to any finite dimension is straightforward. To control for such endogeneity, [Blundell & Powell \(2004\)](#) introduce a control function and assume that

$$u|X, Z \sim u|X, V \sim u|V, \text{ where} \\ V = X^e - m_0(Z), Z = (Z_1, Z_2), \text{ and } E\{V|Z\} = 0. \quad (2)$$

The symbol “ $\sim$ ” denotes “equality in distribution”<sup>1</sup>. The function  $m_0$  is nonparametric. The residual  $V$  is called “control function”, as it is the variable allowing the researcher to control for the presence of endogeneity. Denote with  $G_0(u, v)$  the conditional distribution of  $u$  given  $V = v$  computed at the value  $u$ . The exclusion restriction in Eq. (2) implies that

$$E\{Y|X, Z\} = G_0(X^T \beta_0, V), \quad (3)$$

where  $(X, Z)^T$  denotes the *column* vector gathering all the components of  $X$  and  $Z$  without “repeating” the common ones. Since the cdf of  $u$  is not specified in a parametric way,  $G_0$  is an unknown function.

The above display is the moment condition to be tested in this paper, i.e. the null hypothesis. The test for such a condition is a specification test for the control function model described above. If the null hypothesis is not rejected, the specification of the model cannot be rejected. Conversely, if the null is rejected, the correct specification of the model must be rejected as well<sup>2</sup>.

<sup>1</sup>The condition in Eq. (2) is a conditional independence restriction. It requires that when the residual  $V$  is kept fixed, the error  $u$  is independent from the exogenous variables  $Z$ . This allows to exclude the regressors  $X$  from the conditional distribution of  $u$ .

<sup>2</sup>The vector of parameters parameter  $\beta_0$  can be identified by exclusion restrictions and normalization conditions, as in

*Remark 2.1.* Eq. (3) is implied by the linearity restriction in Eq. (1) and the exclusion restriction in Eq. (2). When the condition in Eq. (3) is rejected, the test does not suggest whether such a rejection is due to the failure of the linearity restriction or the exclusion restriction. Whether the researcher has to reject one or the other depends on the specific application. The control function condition in Eq. (2) can be often justified by economic arguments. Conversely, the linearity in Eq. (1) is often imposed for the aim of simplicity but does not have a strong economic justification. In such cases, a rejection of the null hypothesis would mean a rejection of the linearity restriction in Eq. (1).

## 2.2 Construction of the Test

The goal is to test the hypothesis

$$H_0 : E\{Y | X, Z\} = G_0(X^T \beta_0, X^e - m_0(Z)) \quad \text{versus} \quad H_1 : H_0^c$$

where  $H_0^c$  denotes the logical complement of  $H_0$ . The test described here is an omnibus specification test. Denote with  $f$  the density function of the vector  $(X^T \beta_0, V)$ . For the ease of notation, we introduce the error term

$$\varepsilon := Y - G_0(X \beta_0, V). \quad (4)$$

The null hypothesis  $H_0$  is equivalent to

$$H_0 : E\{\varepsilon \cdot f(X \beta_0, V) | X, Z\} = 0 \quad P\text{-a.s. for some } \beta_0 \in \mathbb{R}^{dim(X)}, \quad (5)$$

where “a.s.” stands for “almost surely”. We introduce the density  $f$  in the above expression for technical reasons, to avoid a random denominator and obtain clean proofs.

The first step to build a test is to transform the above *conditional* moment into a *continuum* of *unconditional* moments. Such a transformation enables to construct the test statistic without estimating  $E\{Y | X, Z\}$ . To this end, let us define the linear operators  $P$  and  $P_n$  as  $Pg = \int g(y, x, z) P(dy, x, z)$  and  $P_n g = (1/n) \sum_{i=1}^n g(Y_i, X_i, Z_i)$ , respectively<sup>3</sup>. By the results in [Bierens \(1982\)](#) and [Stinchcombe & White \(1998\)](#), the hypotheses  $H_0$  and  $H_1$  are equivalent to

$$H_0 : P \varepsilon f \varphi_t = 0 \quad \forall t \in T, \quad H_1 : P \varepsilon f \varphi_t \neq 0 \text{ for almost all } t \in T,$$

where  $T$  is a compact subset of  $\mathbb{R}^{dim(X, Z)}$  encompassing the origin,  $\varphi_t(\cdot) := \gamma(t \cdot)$ , and  $\gamma : \mathbb{R} \rightarrow \mathbb{C}$  is a univariate analytic non-polynomial function. This means that  $\gamma$  is infinitely times continuously

---

[Blundell & Powell \(2004\)](#) and [Rothe \(2009\)](#). This means that some of the exogenous variables, i.e. those in  $Z_2$ , must not appear in the original equation and cannot be part of  $X$ . Conversely, [Escanciano et al. \(2016\)](#) show that in the presence of nonlinearities in the function  $m_0$ , these exclusion restrictions are not necessary, and identify the components of the model without using “instruments”.

<sup>3</sup>Notice that if  $g$  is a nonrandom and deterministic function,  $Pg = Eg(Y, X, Z)$ .

---

**Algorithm 1** Computation of  $S_n$ 

---

Step 1. Regress (nonparametrically)  $X^e$  onto  $Z$  to get an estimate  $\hat{m}_0$  of  $m_0$ , and set  $\hat{V} = X^e - \hat{m}_0(Z)$ .

Step 2. Estimate  $\hat{\beta}$  by minimizing a Semiparametric Least Square criterion.

Step 3. Regress nonparametrically  $Y$  onto  $(X^T \hat{\beta}, \hat{V})$  to get an estimate  $\hat{G}$  of  $G_0$ . Compute the residuals  $\hat{\varepsilon}_i = Y_i - \hat{G}(\hat{X}_i^T \hat{\beta}, \hat{V}_i)$  for  $i = 1, \dots, n$ . Estimate  $f$  by constructing the kernel density estimator of  $(X^T \hat{\beta}, \hat{V})$ . Given  $\hat{\varepsilon}$  and  $\hat{f}$ , obtain the statistic  $S_n$  as in Eq. (6)

---

differentiable and does not have a polynomial form.  $\exp(\cdot)$ ,  $\exp(\cdot/\sqrt{-1})$ ,  $\sin(\cdot)$ ,  $\cos(\cdot)$  are examples of such a function. Let  $(\hat{m}_0, \hat{\beta})$  be estimators of  $(m_0, \beta_0)$ .  $\hat{V} = X^e - \hat{m}_0(Z)$  is an estimator of  $V$ . Denote with  $\hat{\varepsilon} := Y - \hat{G}(X^T \hat{\beta}, \hat{V})$  and  $\hat{f}$  the estimators of  $\varepsilon$  and  $f$ , respectively. A feasible statistic for the test will be

$$S_n = \int |\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{f} \varphi_t \hat{t}_n|^2 \mu(dt), \quad (6)$$

where  $\hat{t}_n$  denotes a trimming function and  $\mu$  is a measure absolutely continuous with respect to the Lebesgue measure<sup>4</sup>. If  $\varphi_t(\cdot) = \exp(it \cdot)$  in Eq. (4), with  $i = \sqrt{-1}$ , the integral in  $S_n$  has a simple closed-form expression,

$$S_n = \frac{1}{n} \sum_{i,j} \hat{\varepsilon}_i \hat{f}_i \hat{\varepsilon}_j \hat{f}_j \cdot \phi_\mu(X_i - X_j, Z_i - Z_j),$$

where  $\phi_\mu$  is the characteristic function of  $\mu$ .

## 2.3 Estimation and Asymptotic Behavior

To test the model, both the regression function  $G(\cdot) = E\{Y | (\beta_0^T X, V) = \cdot\}$  and the density  $f$  need to be estimated. Since the regressors  $(\beta_0^T X, V)$  are not observed, it is natural to proceed to a “three-stage” estimation. First, we get the estimates  $\hat{m}_0$  of  $m_0$ , and hence an estimate  $\hat{V}$  of  $V$ . Second, we compute  $\hat{\beta}$  by a Semiparametric Least Squares criterion, see [Ichimura \(1993\)](#), [Ichimura & Lee \(2010\)](#), and [Escanciano et al. \(2016\)](#). Third, we construct the estimators for  $(G_0, f)$  by replacing the unobserved regressors  $(\beta_0^T X, V)$  with their estimates  $(\hat{\beta}^T \hat{X}, \hat{V})$ . These steps are summarized in [Algorithm 1](#).

Notice that under the null hypothesis,  $\sqrt{n} \mathbf{P}_n \varepsilon f \varphi_t \xrightarrow{d} \mathbf{N}(0, \cdot)$  for any fixed  $t$ , where  $\mathbf{N}(0, \cdot)$  is a normal distribution with a certain variance. Since one expects that  $\hat{\varepsilon} \hat{f} \varphi_t \approx \varepsilon f \varphi_t$ , it should also hold that  $\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{f} \varphi_t \xrightarrow{d} \mathbf{N}(0, \cdot)$ . Hence,  $\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{f} \varphi_t$  will be bounded in probability and so will  $S_n$ . Conversely, under the alternative  $H_1$ , for any fixed  $t$ ,  $\sqrt{n} \mathbf{P}_n \varepsilon f \varphi_t \rightarrow +/\infty$ . So  $\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{f} \varphi_t \rightarrow +/\infty$ , and the statistic  $S_n$  will explode. Thus, for a suitably chosen critical value  $c_{1-\alpha}$ , a test at the  $\alpha$  significance level

---

<sup>4</sup>The trimming  $\hat{t}_n$  excludes from the computation of the statistic those observations on the tails of the distribution. It will be defined more precisely in Section 3.

can be defined as follows:

$$\text{Reject } H_0 \text{ if } S_n > c_{1-\alpha}.$$

Ideally, we would set the critical value  $c_{1-\alpha}$  to the  $1 - \alpha$  quantile of the null distribution of  $S_n$ , but since such a quantile is unknown it must be estimated.

## 2.4 The Bootstrap Test

To obtain the critical value  $c_{1-\alpha}$  we propose a *Wild Bootstrap* procedure. To this end, we start from the wild bootstrap scheme proposed in [Davidson & MacKinnon \(2010\)](#) for the linear two-stages least squares regression. We extend such a scheme to a semiparametric context with generated regressors. The procedure proposed relies on the functional forms imposed under the null hypothesis. Intuitively, since the null hypothesis is imposed in the bootstrap resampling, we should obtain a good approximation of the critical values when the null hypothesis holds true, as the information in  $H_0$  is used. In this context, to obtain a wild-bootstrap procedure leading to a consistent test, it is necessary to mimic that in the “real world” the regressor  $V$  is not observed but has to be estimated in a preliminary step.

Let us define  $\{\xi_i\}_{i=1}^n$  to be a sequence of iid weights independent from the sample data, randomly drawn from a known distribution  $P^\xi$ . For instance,  $P^\xi$  can be set to the standard normal distribution or to any other distribution with mean zero and variance one. Let

$$\begin{aligned} Y_i^* &:= G(\hat{\beta}^T X_i, \hat{V}_i) + \xi_i \cdot \hat{\varepsilon}_i \\ X_i^{e,*} &:= \hat{m}(Z) + \xi_i \cdot \hat{V}_i \end{aligned} \tag{7}$$

If the regressors  $V$  were observed, only the first line of the above display would be sufficient to build a valid bootstrap procedure. Because  $V$  must be estimated in a preliminary step, the introduction of the second line is necessary for a correct bootstrap inference.

By resampling the observations from the above DGP, it is possible to compute the bootstrap version  $S_n^*$  of the statistic  $S_n$ . The steps needed in the bootstrap computations are detailed in [Algorithm 2](#). Let us notice two features. First,  $\hat{V}^*$  is defined as the difference between  $X^e$  and  $\hat{m}_0$  and not as the difference between  $X^{e,*}$  and  $\hat{m}_0(Z)$ . This is because the generated regressor  $V$  has to be bootstrapped only to replicate the estimation error from its unobservability. Such estimation error is generated only by  $\hat{m}_0$ , so it is only  $\hat{m}_0$  that must be bootstrapped. Second, the bootstrap statistic  $S_n^*$  is not “recentered” because the null hypothesis is already imposed in the bootstrap DGP of Eq. (7). More details will be provided in [Section 5](#).

To compute the critical value, one generates  $B$  samples from the bootstrap DGP and gets the collection  $S_{n,1}^*, \dots, S_{n,B}^*$ . The  $(1 - \alpha)$  quantile of such a collection gives an approximation of the critical value for running the test at the  $\alpha$  significance level.

---

**Algorithm 2** Computation of  $S_n^*$ 

---

*Step 1.* Regress nonparametrically  $X^{e,*}$  onto  $Z$  to get an estimator  $\hat{m}_0^*$  and set  $\hat{V}^* = X^e - \hat{m}_0^*(Z)$ . *Step 2.* Compute  $\hat{\beta}^*$  by minimizing a Semiparametric Least Square criterion based on the bootstrap data. *Step 3.* Regress nonparametrically  $Y^*$  onto  $(\hat{\beta}^{*T}X, \hat{V}^*)$  to get  $\hat{G}^*$  (the bootstrap version of  $G$ ). Compute the residuals  $\varepsilon_i^* = Y_i^* - \hat{G}^*(\hat{\beta}^{*T}X_i, \hat{V}_i^*)$  for  $i = 1, \dots, n$ . Compute the bootstrap version of  $f$  by constructing the kernel density estimator of  $(\hat{\beta}^{*T}X, \hat{V}^*)$ , and denote it by  $\hat{f}^*$ . Given  $\hat{\varepsilon}^*$  and  $\hat{f}^*$ , obtain the statistic  $S_n^*$  corresponding to a single bootstrap iteration.

---

### 3 The Estimators

This section introduces in detail the estimators used. From [Algorithm 1](#), the first and third step for constructing the test statistic involve nonparametric estimations. To provide tractable proofs, we use kernel methods. Since nonparametric estimators are biased, several bias terms will appear in the expansion -i.e. the influence function representation- of the empirical process  $\sqrt{n}P_n \varepsilon^* \hat{f} \phi_t$  at the basis of  $S_n$ . To make such bias terms negligible, the approach usually taken in specification testing and semiparametric estimation consists in undersmoothing and employing high-order kernels (see e.g. [Delgado & Manteiga, 2001](#)). In this paper we take a different approach and introduce a bias correction for each nonparametric estimator. The employment of bias corrections is first proposed in specification testing by [Xia et al. \(2004\)](#) who provide a test for a single-index model where all the variables are observed. In this paper, we extend their approach to a context with nonparametrically generated covariates. This appears to be a novel contribution for conducting inference in semiparametric models with generated regressors.

As we highlight in [Section 4](#), the use of bias corrections will have three main advantages. First, the test can be implemented without undersmoothing. Second, the set of bandwidths admissible for the test is larger than the set admissible without the bias correction. Third, relatively low order kernel can be used.

**First-Step Estimation.**  $m_0$  is estimated as

$$\hat{m}_0(z) = \frac{\sum_{i=1}^n X_i^e K_0 \left( \frac{Z_i - z}{h_0} \right)}{\sum_{i=1}^n K_0 \left( \frac{Z_i - z}{h_0} \right)}, \quad (8)$$

where  $h_0$  denotes a bandwidth rate converging to zero and  $K_0$  is a kernel function. Let  $\hat{m}_{0,i} := \hat{m}_0(Z_i)$ . The bias term  $B_0$  of the nonparametric estimators  $\hat{m}_0$  is defined as

$$B_0 := E\{\hat{m}_0(z)\} - m_0(z).$$

We estimate such bias as

$$\hat{B}_0(z) := \mathbb{E}\{\hat{m}_0(z)\} - \hat{m}_0(z),$$

where

$$\begin{aligned} \mathbb{E}\{\hat{m}_0(z)\} &:= \frac{\sum_{i=1}^n \hat{m}_{0,i} \hat{t}_i K_0 \left(\frac{Z_i - z}{h}\right)}{\sum_{i=1}^n K_0 \left(\frac{Z_i - z}{h_0}\right)}, \\ \hat{t}_i &:= 1 \{ \hat{f}_{(X,Z)}(X_i, Z_i) \geq \tau_n \}, \hat{f}_{(X,Z)}(x, z) := \frac{1}{n b^q} \sum_{i=1}^n K_{(X,Z)} \left( \frac{(X, Z)_i - (x, z)}{b} \right). \end{aligned}$$

$b$  is a sequence of bandwidths converging to zero,  $K_{X,Z}$  is a kernel function,  $q = \dim(X, Z)$ ,  $\tau_n$  is a sequence of numbers converging to zero whose features are specified below, while  $\hat{t}$  denotes a trimming whose role is to take care of the random denominators in  $\hat{m}_0$ . Essentially,  $\hat{t}$  excludes from the computation of  $\hat{B}_0$  those observations for which the denominator of  $\hat{m}_0(Z_i)$  is asymptotically close to zero. Such a trimming sequence is introduced for theoretical reasons, but it is not strictly relevant in practice. The intuition about the construction of the estimator  $\hat{B}_0$  is provided at the end of this section. The bias-corrected estimators of  $m_0$  writes as

$$\tilde{m}_0(z) := \hat{m}_0(z) - \hat{B}_0(z). \quad (9)$$

The bias-corrected estimator of  $V$  is given by  $\tilde{V} := X^e - \tilde{m}_0(Z)$ .

**Second-Step Estimation.** For a generic  $\beta \in \mathbb{R}^{\dim(X)}$ , the function  $G_\beta(\cdot) := \mathbb{E}\{Y | (\beta^T X, V) = \cdot\}$  is also estimated by kernel methods, and each unobserved regressor is replaced by its bias-corrected estimate. So

$$\hat{G}_{(\beta, \tilde{V})}(w) := \frac{\sum_{i=1}^n Y_i K \left( \frac{(\beta^T X_i, \tilde{V}_i) - w}{h} \right) \hat{t}_i}{\sum_{i=1}^n K \left( \frac{(\beta^T X_i, \tilde{V}_i) - w}{h} \right) \hat{t}_i}, \quad (10)$$

where  $h$  is a bandwidth parameter and  $K$  is a kernel function. Similarly to the first-step estimation, the trimming  $\hat{t}$  is introduced in  $\hat{G}_{(\beta, \tilde{V})}$  to take care of the random denominator in  $\tilde{m}_0$ . To estimate the vector  $\beta_0$ , a Semiparametric Least-Square criterion is minimized ([Ichimura, 1993](#); [Ichimura & Lee, 2010](#); [Escanciano et al., 2016](#)),

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbf{B}} \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{G}_{(\beta, \tilde{V})}(\beta^T X_i, \tilde{V}_i)]^2 \hat{t}_i \quad (11)$$

where  $\mathbf{B}$  is a compact set containing the true parameter  $\beta_0$ . Notice that  $\hat{G}_{(\beta, \tilde{V})}$  does not contain a bias correction. The reason is technical and will be explained in [Section 3](#).

**Third-Step Estimation.** An estimator of  $G(\cdot) = E\{Y | (\beta^T X, V) = \cdot\}$  can be obtained by regressing nonparametrically  $Y$  onto  $(\beta^T X, \tilde{V})$ . It is computed by replacing  $\beta$  with  $\hat{\beta}$  in Eq. (10). Denote  $\hat{G}_{(\hat{\beta}, \tilde{V}), i} := \hat{G}_{(\hat{\beta}, \tilde{V})}(\hat{\beta}^T X_i, \tilde{V}_i)$ . The bias of  $\hat{G}_{(\hat{\beta}, \tilde{V})}$  and the density  $f$  are respectively estimated as

$$\hat{B}_{\tilde{V}}(w) := E\{\hat{G}_{(\hat{\beta}, \tilde{V})}(w)\} - \hat{G}_{(\hat{\beta}, \tilde{V})}(w) \quad (12)$$

with

$$\hat{E}\{\hat{G}_{(\hat{\beta}, \tilde{V})}(w)\} := \frac{\sum_{i=1}^n \hat{t}_{(\hat{\beta}, \tilde{V}), i} \cdot \frac{(\hat{\beta}^T X_i, \tilde{V}_i) - w}{h}}{\sum_{i=1}^n \hat{t}_i \cdot K \cdot \frac{(\hat{\beta}^T X_i, \tilde{V}_i) - w}{h}},$$

and

$$\hat{f}(w) = \frac{1}{nh^d} \sum_{i=1}^n K \cdot \frac{(\hat{\beta}^T X_i, \tilde{V}_i) - w}{h} \hat{t}_i.$$

The trimming  $\hat{t}$  is introduced in  $E\{\hat{G}_{(\hat{\beta}, \tilde{V})}(w)\}$  to control for the random denominators in  $\hat{G}_{(\hat{\beta}, \tilde{V})}$  and  $\tilde{V}$ . The bias-corrected estimator of  $G_0$  writes as

$$\tilde{G}_{(\hat{\beta}, \tilde{V})}(w) := \hat{G}_{(\hat{\beta}, \tilde{V})}(w) - \hat{B}_{\tilde{V}}(w).$$

At this stage, we have introduced all the elements required to compute  $S_n$ . So,

$$\tilde{\varepsilon} := Y - \tilde{G}_{(\hat{\beta}, \tilde{V})}(\hat{\beta}^T X, \tilde{V}) \text{ and } S_n := \int \left| \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \tilde{\varepsilon} \hat{f}_n \right|_H^2(dt).$$

*Remark 3.1.* The method presented in this section for computing the residuals is different from the one in [Xia et al. \(2004\)](#). They consider a semiparametric environment and do not have a problem of non-observability of the regressors. Since they only need to perform a single nonparametric estimation, they introduce a single bias correction. Conversely, in this context we have to deal with a semiparametric model containing nonparametrically generated covariates. So we introduce two bias corrections: one for the estimation of  $V$  and another for the estimation of  $G_0$ .

**Intuition and interpretation of the bias correction.** Consider the bias correction applied to  $\hat{m}_0$ , the nonparametric estimator of  $m_0(\cdot) = E\{X^e | Z = \cdot\}$ . The bias of  $\hat{m}_0$  is formally defined as  $B_0(\cdot) = E\{\hat{m}_0(\cdot)\} - m_0$ . Its estimate in Eq. (9) can be interpreted as a Wild Bootstrap estimate ([Xia et al., 2004](#)). To explore this point, consider the bootstrap DGP

$$X_i^{e,*} = \hat{m}_0(Z_i) + \xi_i \cdot \hat{V}_i \text{ with } V_i = X_i - \frac{e}{i} \hat{m}(Z_0),$$

where  $\{\xi_i\}_{i=1}^n$  is a sequence of bootstrap weights independent from the sample data, with mean zero and variance one. Given the sample  $\{X_i^{e,*}, Z_i\}_{i=1}^n$ , we can compute the bootstrap version of  $\hat{m}_0$ . Let  $\hat{m}_0^*$  denote the kernel estimator computed with the bootstrap data. In the bootstrap context the sample data is fixed and only the bootstrap weights  $\{\xi_i\}_{i=1}^n$  are random. So, the bias of  $\hat{m}^*$  is defined as  $Bias\{\hat{m}^*_0(z)\} = E^\xi\{\hat{m}^*_0(z)\} - \hat{m}_0(z)$ , where  $E^\xi$  is the expectation that considers only the bootstrap weights  $\{\xi_i\}_{i=1}^n$  as random and the sample data as fixed. It is easy to show that

$$Bias\{\hat{m}^*_0(z)\} = E^\xi\{\hat{m}^*_0(z)\} - \hat{m}_0(z) = E\{\hat{m}_0(z)\} - \hat{m}_0(z) = Bias\{\hat{m}_0(z)\}.$$

By the preceding display, the bias correction used in this paper can be interpreted as a bootstrap estimate of the bias term. A similar reasoning can be applied to the bias correction of  $\hat{G}_{(\beta, V)}^*$ <sup>5</sup>.

## 4 Asymptotic Analysis

In this section we provide the assumptions and present the asymptotic behavior of the statistic. Define  $G_\beta(\cdot) := E\{Y | (\beta^T X, V) = \cdot\}$  and let  $f_\beta$  be the density of  $(\beta^T X, V)$ , where  $\beta$  ranges over a compact set  $\mathbf{B}$  encompassing  $\beta_0$ . Notice that  $f_{\beta_0} = f$  and  $G_{\beta_0} = G_0$ . Recall that  $(X, Z)$  stands for the vector gathering the components of  $X$  and  $Z$  without repeating the common ones. Also, let  $f_0$  be the density of  $Z$ . For any natural number  $c$ , we denote with  $c^\dagger$  the largest even number weakly smaller than  $c$ <sup>6</sup>. For a vector  $A = (A_1, \dots, A_S)$  of natural numbers,  $A = A_1 + \dots + A_S$  and the differential operator  $\partial^A$  is defined as

$$\partial_u^A g(u_1, \dots, u_S) = \frac{\partial^A}{\partial^{A_1} u_1 \dots \partial^{A_S} u_S} g(u_1, \dots, u_S).$$

The following definitions are helpful for the presentation of the assumptions at the basis of our test. We first introduce a class of kernel functions.

**Definition 4.1.**  $\mathcal{K}_r^\lambda$  is the class of mappings  $(u_1, \dots, u_S) \mapsto \prod_{s=1}^S k(u_s)$ , where  $k$  is a univariate kernel of order  $r$  that has a bounded support and is  $\lambda$  times continuously differentiable with bounded derivatives.

The next definition introduces two classes of smooth functions.

**Definition 4.2.**  $\mathcal{G}^\lambda$  is the class of mappings  $(w, \beta) \mapsto g(w, \beta)$  such that: (i) for all  $\beta \in \mathbf{B}$ ,  $w \mapsto g(w, \beta)$  is  $\lambda$  times continuously differentiable with uniformly bounded derivatives,  $\sup_{\beta, w} |\partial_w^A g(w, \beta)| < \infty$  for all  $A \leq \lambda$ ; (ii)  $\beta \mapsto g(w, \beta)$  is continuously differentiable.

<sup>5</sup>The bias correction adopted in this work would coincide with the twicing kernel method in the presence of a fixed design, see [Newey et al. \(2004\)](#)

<sup>6</sup>If  $c$  is even,  $c^\dagger = c$ , while if  $c$  is odd  $c^\dagger = c - 1$ .



$\mathbf{E}_\kappa^r$  is the class of mappings  $(x, z, w, \beta) \mapsto \gamma(x, z, w, \beta)$  such that: (i)  $\gamma$  is  $\lambda$  times continuously differentiable in  $w$  with  $\sup_{\beta, (x, z), w} |\partial_w^A \gamma(x, z, w, \beta)| < \infty$  for all  $A \leq \lambda$ ; (ii)  $\gamma$  is  $\kappa$  times continuously differentiable in  $(x, z)$  with  $\sup_{\beta, (x, z), w} |\partial_{(x, z)}^A \gamma(x, z, w, \beta)| < \infty$  for all  $A \leq \kappa$ ; (iii)  $\gamma$  is differentiable in  $\beta$  with uniformly bounded derivatives.

**Assumption 1.**  $\{Y_i, X_i, Z_i\}_{i=1}^n$  is an IID sequence of bounded random variables defined over the probability space  $(\Omega, \mathcal{A}, P)$ .

**Assumption 2 (Smoothness).** (i) The mappings  $(w, \beta) \mapsto G_\beta(w)$  and  $(w, \beta) \mapsto f_\beta(w)$  belong to the class  $\mathbf{G}^r$ ; (ii) for all  $\beta \in \mathcal{B}$ ,  $(X, Z)$  admits a density conditionally on  $(\beta^T X, V)$ , denoted as  $(x, z, w) \mapsto f_{(X, Z) | (\beta^T X, V)}(x, z | (\beta^T X, V) = w)$  which belongs to the class  $\mathbf{E}_{r_0}^r$ ; (iii)  $m_0, f_0 \in \mathbf{G}^{r_0}$ .

**Assumption 3 (Kernels)** (i)  $K \in \mathbf{K}_d^{r_0}$ ; (ii)  $K_0 \in \mathbf{K}_{p_0}^{r_0}$ ; (iii)  $K_{X, Z} \in \mathbf{K}_{\dim(X, Z)}^{r_3}$ .

Assumption 1 is common in the literature on nonparametric estimation and testing. Assumption 2 imposes a certain degree of smoothness on the functions involved, which is connected with the orders of the kernels defined in Assumption 3. Such a smoothness condition is common in the literature on semiparametric and nonparametric estimation, and can be considered as a mild one (see e.g. [Delgado & Manteiga, 2001](#); [Escanciano et al., 2016](#); [Neumeyer & Van Keilegom, 2010](#))<sup>7</sup>. The continuity of the variables  $(X, Z)$  is only assumed for simplicity, to provide clean proofs. Our results can be easily extended to the case where  $(X, Z)$  involves some discrete components, assuming the existence of densities with respect to the mixed Lebesgue-counting measure (see [Huang et al., 2016](#)).

Since the framework at hand is featured by nonparametrically generated regressors, we need some conditions on the rates at which the densities of the observed variables go to zero on the tails. So, define

$$p_n := P(f(X | \beta_0, V)) < \tau_n, p_{n,0} := P(f_0(Z) < \tau_n), p_{n,3} := P(f_{(X, Z)}(X, Z) < \frac{3\tau_n}{2}).$$

Denote with  $q$  the dimension of  $(X, Z)$ .

**Assumption 4 (Bandwidth Rates).** (i)  $\frac{\log n}{n h_0^{p_0+2s_0} \tau_n^{2s_0+2}} \rightarrow 0$  for  $s_0 = 1 + \frac{p_0^\dagger}{2}$ ,  $\frac{(\log n)^2}{n h_0^{2p_0} \tau_n^{12}} \rightarrow 0$ ,  $n h_0^{4r_0} \tau_n^{-4} \rightarrow 0$ ;

<sup>7</sup>The existence of the density  $f_{(X, Z) | (X^T \beta, V)}$  is also assumed in [Escanciano et al. \(2014\)](#). Notice, however, that as long as the variable  $(X, Z)$  admits a density -i.e.  $P^{(X, Z)}$  is dominated by the Lebesgue measure-, by the Radon-Nikodym Theorem also  $(X, Z)$  conditionally on  $(\beta^T X, V)$  will admit a density, and the existence of  $f_{(X, Z) | (X^T \beta, V)}$  will be ensured. In this context, therefore, the existence of such a function is not a strong restriction.

$$(ii) \frac{(\log n)^2}{n h^{2d+4} \tau_n^8} \rightarrow 0, n \cdot h^{4r} \cdot \tau_n^{-4} \rightarrow 0;$$

**Assumption 5 (Trimming Sequences).** (i)  $\frac{h_0}{\tau_n} = o(1)$ ,  $\frac{h}{\tau_n} = o(1)$ ,  $p_n = o(l_n^{-1})$ ,  $p_{n,0} = o(l_n^{-1})$ ,  $p_{n,3} = o(l_n^{-1})$ , where  $l_n$  satisfies  $\frac{n^{1/4}}{h^{d+1} \tau_n^2 l_n} \rightarrow 0$ ,  $\frac{n^{1/4}}{h^{p_0} \tau_n^2 l_n} \rightarrow 0$ ,  $\frac{n^{1/2}}{l_n} \rightarrow 0$ ;

$$(ii) b \text{ satisfies: } n \cdot b^{q+5} \cdot \tau_n \rightarrow +\infty, \frac{\log n}{n \cdot b^q \tau_n} \rightarrow 0, \frac{b^3}{\tau_n} \rightarrow 0, \tau_n \cdot l_{n,3} \cdot b^q \rightarrow +\infty;$$

(iii) there exists constants  $\eta > 0$  small enough and  $N$  large enough such that for all  $n$  greater than  $N$ ,  $1\{f_{(X,Z)}(x, z) \geq \tau_n\} \leq \sup_{\beta \in B} 1\{f_{(X^T \beta, V)}(x^T \beta, v) \geq \eta \tau_n\}$  and  $1\{f_{(X,Z)}(x, z) \geq \tau_n\} \leq 1\{f_0(z) \geq \eta \tau_n\}$ .

Assumption 4 is introduced for several reasons. First, it implies that the estimators of the generated regressors  $V$  are  $n^{-1/4}$ -consistent, similarly to [Escanciano et al. \(2016\)](#) or [W.K. Andrews \(1994\)](#). Second, it ensures that the estimators of the generated regressors are asymptotically enclosed in a class of functions that is sufficiently regular, similarly to [Escanciano et al. \(2016\)](#), [Neumeyer & Van Keilegom \(2010\)](#), and [Chen et al. \(2003\)](#). Third, it allows  $\tilde{G}_{(\hat{\beta}, \tilde{V})}$  and  $\partial \tilde{G}_{(\hat{\beta}, \tilde{V})}$  to be  $n^{-1/4}$ -consistent, similarly to [Rothe \(2009\)](#). These  $n^{-1/4}$  convergence rates are required for the presence of the nonparametrically-generated regressors. Intuitively, to handle the estimation error arising from the replacement of  $m_0$  with  $\tilde{m}_0$ , we use a procedure similar to [Rothe \(2009\)](#). We combine a 1st-order Taylor expansion with these  $n^{-1/4}$ -rates, to have an approximation of the type

$$\sqrt{n} \tilde{P}_n \tilde{G}_{(\hat{\beta}, \tilde{V})}(\beta X, \tilde{V}) \approx \sqrt{n} \tilde{P}_n \tilde{G}_{(\hat{\beta}, \tilde{V})}(X\beta, V) + \sqrt{n} \tilde{P}_n \partial G_0(X\beta, V) \cdot ((X\beta, \tilde{V}) - (X\beta, V)).$$

*Remark 4.1.* The condition  $\frac{\log n}{n h_0^{p_0+2s_0} \tau_n^{2s_0+2}} \rightarrow 0$  implies that the estimator of  $m_0$  belongs asymptotically to a class of functions that is sufficiently regular. Conversely, the condition  $\frac{(\log n)^2}{n h_0^{2p_0} \tau_n^{12}} \rightarrow 0$  ensures that the estimator of  $m_0$  is  $n^{-1/4}$ -consistent. These conditions are *both* used only in the derivation of the influence-function representation of  $\sqrt{n} \cdot (\hat{\beta} - \beta_0)$ . If one assumes that such an influence-function representation holds, then it is possible to derive the asymptotic behavior of  $S_n$  without using the condition  $\frac{\log n}{n h_0^{p_0+2s_0} \tau_n^{2s_0+2}} \rightarrow 0$ , and only requiring that  $\frac{(\log n)^2}{n \cdot h_0^{2p_0+12} \cdot \tau_n} \rightarrow 0$ . The proof of the asymptotic behavior of  $S_n$  in such a case would be based on U-Process Theory and not on Asymptotic Stochastic Equicontinuity arguments as in this paper. For an employment of U-Process theory in specification testing and semiparametric estimation, see [Sherman \(1994\)](#), [Delgado & Manteiga \(2001\)](#), [Lavergne et al. \(2015b\)](#).

We also highlight that the  $n^{-1/4}$ -consistency of the first derivative  $\partial \tilde{G}_{(\hat{\beta}, \tilde{V})}$  is employed only to obtain shorter proofs. It can be avoided by dealing with U-Processes of order 7.

The part of Assumption 4 linking the bandwidth  $(h, h_0)$  to the orders of the kernels  $(r, r_0)$  allows to control the bias terms appearing in the expansion of the empirical process  $\sqrt{n}P_n \varepsilon \tilde{f} \hat{\varphi}_t \hat{t}_n$ . A similar condition is used in [Xia et al. \(2004\)](#) in a specification test for a single-index model where all the co-variables are observed. If we ignore the trimming rates  $\tau_n$ , Assumption 4 requires that  $n \cdot h_0^{4r_0} \rightarrow 0$  and  $n \cdot h^{4r} \rightarrow 0$ . This implies that the bandwidths for  $\tilde{m}_0$  and  $\tilde{G}_{(\hat{\varphi}, \tilde{V})}$  can be set to the rate that minimizes the Mean-Squared error of the nonparametric estimators (see [Li & Racine, 2006](#)). These rates are allowed by the introduction of the bias corrections. Without these corrections, the conditions  $n \cdot h_0^{4r_0} \rightarrow 0$  and  $n \cdot h^{4r} \rightarrow 0$  must be replaced by  $n \cdot h_0^{2r_0} \rightarrow 0$  and  $n \cdot h^{2r} \rightarrow 0$ , respectively. After such a replacement the rates minimizing the mean squared errors could not be implemented in the construction of the test statistic. This feature is called undersmoothing and is commonly used in specification testing or in the expansion of nonparametric estimators to deal with the bias terms (see e.g. [Delgado & Manteiga, 2001](#); [Escanciano et al., 2014](#); [Escanciano et al., 2016](#); [Huang et al., 2016](#)). Using the conditions  $n \cdot h_0^{4r_0} \rightarrow 0$  and  $n \cdot h^{4r} \rightarrow 0$  instead of  $n \cdot h_0^{2r_0} \rightarrow 0$  and  $n \cdot h^{2r} \rightarrow 0$  has several advantages. First, as already noticed, they allow to set the bandwidths proportional to the optimal rates that minimize the mean-squared error. Thus, they implicitly provide a selection rule for the smoothing parameters that can be used in the test. A second advantage is that Assumption 4, from a practical standpoint, allows to implement the test by cross-validated bandwidths. The Cross-Validation method, in fact, delivers bandwidth rates that are asymptotically equivalent to the rates minimizing the mean squared error. Third, Assumption 4 widens the spectrum of bandwidths and kernel orders admissible with respect to the case where the bias corrections are not used. In this latter case, the rates  $n \cdot h_0^{2r_0} \rightarrow 0$  and  $n \cdot h^{2r} \rightarrow 0$  are employed, and the kernel orders that should be employed to deal with the bias terms must be relatively larger. Larger kernel orders normally generate irregular behaviors of the statistic in finite samples, as they inflate the variance of the estimates (see e.g. [Rothe, 2009](#); [Jones & Signorini, 1997](#)). To contain such variances, it is advised to employ low-order kernels which however make the bias impact more pronounced. In other words, the presence of the bias corrections allows for an improvement of the bias-variance trade-off compared to the case where the corrections are not present.

*Remark 4.2.* The bias correction for  $\hat{G}_{(\beta, \tilde{V})}$  is not included in the SLS criterion for the estimation of  $\beta_0$  (see Eq. 11). The influence function representation of  $\hat{\beta}$  can still be obtained by imposing  $nh^{4r} = o(1)$  and without employing a bias correction, thanks to the fact that  $E\{\nabla_{\beta} G_{(\beta, V)}(X^T \beta, V) |_{\beta=\beta_0} (X^T \beta_0, V)\} = 0$  (see [Ichimura, 1993](#)). This feature ensures that the bias terms appearing in the influence function representation of  $\hat{\beta}$  are identically zero.

In Assumption 4 and 5, the rates of  $h$  and  $h_0$  are connected with the rate  $\tau_n$  appearing in the trimming. Similar conditions can be found in [Escanciano et al. \(2014\)](#) and [Escanciano et al. \(2016\)](#). These conditions are required because the unobserved regressor  $V$  is replaced with its estimate  $\tilde{V}$  containing

a random denominator we need to control for:

The presence of the rate  $l_n$  in Assumption 5 is due to the introduction of the trimming  $\hat{t}_n$  in  $\tilde{m}_0$  and  $\tilde{G}_\beta$ . We need that such a sequence must converge to zero at a sufficiently fast rate, to avoid a bias coming from the presence of the trimming sequences. However, for the practical implementation of the test, the specification of the trimming rates  $\tau_n$  and  $l_n$  can be avoided.

Assumption 5(iii) is similar to Assumption 7 in Escanciano *et al.* (2014). It has a technical nature and essentially avoids the introduction of multiple trimmings. Specifically, to control for the random denominators present in  $\tilde{m}_0$  and  $\tilde{G}_{(\hat{x}, \hat{v})}$ , we should have introduced three different trimmings, one for each random denominator appearing in the statistic  $S_n$ . By avoiding the introduction of three different trimmings, Assumption 5(iii) dramatically simplifies the proofs<sup>8</sup>.

**Assumption 6 (Pseudo-True Value).** *The mapping  $\beta \mapsto \mathbb{E}\{ (Y - G_\beta(X^T \beta, V))^2 \}$  admits a unique minimum.*

This last assumption imposes the existence of a unique pseudo-true value of the finite dimensional parameter. Such a condition is common to any specification test for nonlinear models where estimation is obtained by nonlinear criteria (see e.g. Bierens, 1982; Lavergne & Patilea, 2013; Escanciano *et al.*, 2018). Assumption 6 plays a double role: under the null hypothesis  $H_0$  it ensures the identification of  $\beta_0$ , while under the alternative  $H_1$  it ensures that the estimator  $\hat{\beta}$  has a well defined limit in probability.

## 4.1 The Asymptotic Behavior of $S_n$

We first define a collection of functions linked with the asymptotic behavior of the process  $\sqrt{n} \tilde{\mathbb{P}}_n \tilde{f} \varphi_t$ . Let

$$\begin{aligned} \phi_{0,t}(y, x, z) := & \sum (y - G_0(x\beta_0, v)) \cdot \psi_{0,t}(x, z) + \\ & + \mathbb{E}' \psi_{0,t}(X, Z) \cdot \partial_2 G_0(X\beta_0, V)^T \cdot Z = z' \cdot (x^e - m_0(z)), \end{aligned} \quad (13)$$

where  $\{ \psi_{0,t} : t \in \mathbb{T} \}$  is a class of weights defined in Appendix A (see Remark 26).

**Proposition 1.** *Let Assumptions 1-6 hold.*

- (i) Under  $H_0$ ,  $\sqrt{n} \tilde{\mathbb{P}}_n \tilde{f} \varphi_t = \sqrt{n} \mathbb{P}_n \phi_{0,t} + o_P(1)$ , uniformly in  $t \in \mathbb{T}$ .
- (ii) Under  $H_0$ ,  $S_n \xrightarrow{d} \int |\mathbf{G}|^2 \mu(dt)$ , where  $\mathbf{G}$  is a Gaussian process defined by the collection of covariances  $\{\Phi(t_1, t_2) = \mathbb{P} \phi_{0,t_1} \phi_{0,t_2} : t_1, t_2 \in \mathbb{T}\}$ .

<sup>8</sup>It is possible to remove Assumption 5(iii) at the cost of longer proofs.  $Z$ ,  $X$ , and  $(X\beta, V)$  are transformations of the random variable  $(X, Z)$ . Assumption 5(iii) imposes that when the density of  $(X, Z)$  is larger than a certain small value, asymptotically also the densities of these transformations must be larger than a small value. If all the components of  $(X, Z)$  were discrete such an assumption would automatically hold.

(iii) Under  $H_1$ ,  $\frac{S_n}{n} \rightarrow^P c$  with  $c > 0$ .

The process  $\{\sqrt{n}P_n\phi_t : t \in T\}$  defines the influence function representation of  $\{\sqrt{n}P_n\tilde{\varepsilon}\hat{f}_{\phi_t} : t \in T\}$ , or in other words its Bahadur's expansion. It hence determines the asymptotic behavior of the statistic. A similar expansion is obtained in [Escanciano et al. \(2014\)](#), although under different conditions on the bandwidths and without using a bias correction. This shows that the introduction of the bias correction does not have any impact on the influence function representation of  $\{\sqrt{n}P_n\tilde{\varepsilon}\hat{f}_{\phi_t} : t \in T\}$ , and the same expansion could be obtained by undersmoothing, i.e. by imposing that  $nh^{2r} = o(1)$  and  $nh_0^{2r_0} = o(1)$ . From Eq. (13) we notice that the expansion of the process  $\{\sqrt{n}P_n\tilde{\varepsilon}\hat{f}_{\phi_t} : t \in T\}$  not only contains a weighted sum of the error  $\varepsilon$ , the population counterpart of the residuals  $\tilde{\varepsilon}$ , but also a weighted sum of the control function  $V$ . The appearance of the latter term can be considered as the “price to pay” for not observing  $V$  and having to estimate it in a preliminary step, similarly to [Hahn & Ridder \(2013\)](#). The weights attached to  $V$  depend on the sensitivity of  $G_0$  with respect to such unobserved regressor. If  $V$  is not significant in  $G_0$ , i.e. if  $\partial_2 G_0$  is identically zero, the influence function representation of  $\{\sqrt{n}P_n\tilde{\varepsilon}\hat{f}_{\phi_t} : t \in T\}$  would be the same as if the errors  $\varepsilon$  were estimated observing  $V$ .

To compute the critical value, several methods proposed in the literature on specification testing could be adapted to our case. [Bierens & Ploberger \(1997\)](#) obtain the critical values of a specification test for a parametric regression by estimating an upper bound of the quantiles of the distribution  $|\tilde{G}|_{\mu}^2(dt)$ , where  $\tilde{G}$  is a Gaussian process. As a consequence, such a method will be conservative and may have a low power. [Horowitz \(2006\)](#) approximates the asymptotic critical values by estimating the eigenvalues of the covariance matrix operator which characterize the process  $G$ . Since the number of such eigenvalues is infinite (see [Horowitz, 2006](#); [Bierens & Ploberger, 1997](#)), such a procedure requires an arbitrary cut. Moreover, since in our context the covariance matrix operator of  $G$  contains nonparametric derivatives (see the expression of  $\Phi$  in Proposition 4.1 and Eq. 13), the estimation of these eigenvalues requires the estimates of nonparametric derivatives which generally have low convergence rates. Such low convergence rates might compromise the capacity of the test to provide a reliable inference in finite samples. Finally, [Delgado & Manteiga \(2001\)](#) and [Xia et al. \(2004\)](#) provide a specification test where all the variables are observed, and propose to obtain the critical value by a wild-bootstrap procedure. In the next section we use a similar approach.

## 5 The Bootstrap Test

To construct a Wild Bootstrap test, we start from the wild bootstrap procedure proposed in [Davidson & MacKinnon \(2010\)](#) for the linear two-stages least-squares model. We extend such a method to a semiparametric context with nonparametrically generated regressors, where the estimators of such variables contain bias corrections. Our bootstrap test can also be seen as an extension to a context with generated regressors of the wild bootstrap proposed in [Delgado & Manteiga \(2001\)](#) and [Xia et al. \(2004\)](#).

While in these papers all the regressors are observed, our method provides a novel way to conduct inference in semiparametric models with generated regressors.

Let us define  $\{\zeta_i\}_{i=1}^n$  to be a sequence of weights independent from the sample data, with mean zero and variance equal to one. Let

$$\begin{aligned} Y_i^* &:= \hat{G}_{(\hat{\beta}, \hat{V})}(\hat{\beta} X_i, \hat{V}_i) + \zeta_i \cdot \hat{\varepsilon}_i \text{ with } \hat{\varepsilon}_i = Y_i - \hat{G}_{(\hat{\beta}, \hat{V})}(\hat{\beta} X_i, \hat{V}_i) \\ X_i^{e,*} &:= \hat{m}(Z_i) + \zeta_i \cdot \hat{u}_i \text{ with } \hat{u}_i = X_i - \hat{m}(Z_i) \end{aligned}$$

for  $i = 1, \dots, n$ . We notice three features of the above bootstrap DGP. First, the estimators displayed do not involve bias corrections. Such corrections are not necessary in the bootstrap DGP for obtaining the validity of the bootstrap inference. Second, the bootstrap weights  $\{\zeta_i\}_{i=1}^n$  have to be the same in the first and the second equation. This preserves the covariance structure of the error terms compared to the original sample, ensuring the validity of the bootstrap inference. Third, the introduction of the second line for  $X_i^{e,*}$  is due to the non-observability of the regressors  $V$ . This allows us to mimic the first-step estimation that must be performed in the original sample.

**The bootstrap counterpart of  $S_n$**  Given  $n$  observations  $\{Y_i^*, X_i^{e,*}\}_{i=1}^n$  generated from the bootstrap DGP, the bootstrap counterpart  $S_n^*$  of  $S_n$  can be computed according to the following steps.

*First-Step Estimation for Bootstrap.* Define

$$\hat{m}_0^*(z) = \frac{\sum_{i=1}^n X_i^{e,*} K_0 \left( \frac{Z_i - z}{h_0} \right) \hat{t}_i}{\sum_{i=1}^n K_0 \left( \frac{Z_i - z}{h_0} \right)} \quad (14)$$

Given the bias estimate  $\hat{B}_0(\cdot)$  from Section 3, the bootstrap counterpart of  $\tilde{m}_0$  writes as

$$\tilde{m}_0^*(z) := \hat{m}_0^*(z) - \hat{B}_0(z).$$

*Remark 5.1.* The bias correction used for  $\tilde{m}_0^*$  is the same as the one for  $\tilde{m}_0$  and is not computed according to the bootstrap data. As discussed in Section 3, such a bias correction corresponds to the *true bias* in the bootstrap context.

*Second-Step Estimation for Bootstrap.* Set  $\tilde{V}^* := X^e - \tilde{m}_0^*(Z)$ . The bootstrap counterpart of  $\hat{G}_{(\beta, \tilde{V})}$  is

$$\hat{G}_{(\beta, \tilde{V}^*)}^*(w) := \frac{\sum_{i=1}^n Y_i^* K \left( \frac{(\beta X_i^T, \tilde{V}^*)^* - w}{h} \right)}{\sum_{i=1}^n K \left( \frac{(\beta X_i^T, \tilde{V}^*)^* - w}{h} \right)} \cdot \frac{\sum_{i=1}^n \hat{t}_i}{\sum_{i=1}^n \hat{t}_i}.$$

The bootstrap version of the estimator  $\hat{\beta}$  is given by

$$\hat{\beta}^* = \operatorname{argmin}_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^n [Y_i^* - \hat{G}_{(\beta, \tilde{V}^*)}^*(\beta^T X_i, \tilde{V}_i)]^2 \cdot \hat{t}_i.$$

*Remark 5.2.* First, we notice that  $\tilde{V}^* = X^e - \tilde{m}_0^*(Z)$ , so the bootstrapped control function is *not* given by the difference  $X^{e,*} - \tilde{m}_0^*(Z)$ . Second, the vector  $X$  includes  $X^e$  and does *not* include  $X^{e,*}$ . This is because the non-observability of the regressor  $V$  at the sample level is only due to the non-observability of  $m_0$ . Hence, to replicate the estimation error from the non-observability of  $V$ , we need to bootstrap only  $\tilde{m}_0$ .

*Third-Step Estimation for Bootstrap.* Let  $\hat{G}_{(\hat{\beta}, \hat{V}), i} := \hat{G}_{(\hat{\beta}, \hat{V})}(\hat{\beta}^T X_i, \hat{V}_i)$ . Notice that  $\hat{G}_{(\hat{\beta}, \hat{V})}$  does not involve bias corrections. Define

$$\hat{B}_{\hat{V}}(w) := \mathbb{E}\{\hat{G}_{(\hat{\beta}, \hat{V})}(w)\} - \hat{G}_{(\hat{\beta}, \hat{V})}(w),$$

with

$$\mathbb{E}\{\hat{G}_{(\hat{\beta}, \hat{V})}(w)\} = \frac{\sum_{i=1}^n \frac{\hat{t}_i K \left( \frac{(\hat{\beta}^T X_i, \hat{V}_i) - w}{h} \right)}{\sum_{i=1}^n \hat{t}_i K \left( \frac{(\hat{\beta}^T X_i, \hat{V}_i) - w}{h} \right)}}{\sum_{i=1}^n \hat{t}_i K \left( \frac{(\hat{\beta}^T X_i, \hat{V}_i) - w}{h} \right)}.$$

The bootstrap counterpart of  $\hat{f}$  writes as

$$\hat{f}^*(w) := \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{(\hat{\beta}^T X_i, \tilde{V}_i) - w}{h} \right) \cdot \hat{t}_{n,i}.$$

The bias correction  $\hat{B}_{\hat{V}}$  is computed with the sample data and *not* with the bootstrapped data. We also notice that such a bias correction is based on the estimator  $\hat{V}$  and not on  $\tilde{V}$ . The bootstrap counterpart of  $\tilde{G}_{\tilde{V}(\hat{\beta})}$  writes as

$$\tilde{G}_{(\hat{\beta}, \hat{V})}^*(w) := \hat{G}_{(\hat{\beta}, \hat{V})}^*(w) - \hat{B}_{\hat{V}}^*(w).$$

*Remark 5.3.* The bias correction  $\hat{B}_{\hat{V}}$  used in the bootstrap context corresponds to the true bias of the

kernel smoothing of  $Y^*$  onto  $(X^T \hat{\beta}, \hat{V})$ . The reasoning is the same as in Section 3. The original bias correction  $\hat{B}_{\hat{V}}$  used at the sample level is not implemented in the bootstrap estimators, as it does not correspond to the *true* bias of the kernel estimator in the bootstrap context.

*The Bootstrapped Statistic.* Define the bootstrap version of the residuals  $\tilde{\varepsilon}$  as

$$\tilde{\varepsilon}^* = Y^* - \tilde{G}_{(\hat{\beta}^*, \tilde{V}^*)}^*(X\hat{\beta}^*, \tilde{V}^*).$$

The bootstrap version of the statistic writes as

$$S_n^* := \int |P_{t, \tilde{\varepsilon}^*}^* \hat{f}^* \hat{t}_n|^2 \mu(dt).$$

*Remark 5.4.* The bootstrap statistic  $S_n^*$  does not include a “recentering” term, as the null  $H_0$  is imposed in the bootstrap DGP, so we are already bootstrapping under the null.

**Consistency of the Bootstrap Test.** The estimate of the  $(1 - \alpha)$ -quantile of the null distribution of  $S_n$  is defined as

$$\hat{c}_{1-\alpha} := \inf^c c : P_n^{\tilde{\zeta}}(S_n^* \leq c) \geq 1 - \alpha',$$

where  $P_n^{\tilde{\zeta}} = \otimes_{i=1}^n P_n^{\tilde{\zeta}_i}$  denotes the probability measure that considers as random the weights  $\{\tilde{\zeta}_i\}_{i=1}^n$  and as fixed the sample data. According to the decision rule described in Section 2, we reject the null  $H_0$  at the  $\alpha$  significance level as long as  $S_n > \hat{c}_{1-\alpha}$ . Let us denote with  $P$  the joint probability measure resulting from the product between the two measures  $P$  and  $P$ , i.e.  $P_n^{\tilde{\zeta}} = P \otimes P$ . The following proposition shows the validity of the wild-bootstrap scheme.

**Proposition 2.** *Let Assumptions 1-6 hold.*

- (i) *Under  $H_0$ ,  $\sqrt{n}P_n \tilde{\varepsilon}^* \hat{f}^* \varphi_t = \sqrt{n}P_n^{\tilde{\zeta}} \phi_{0,t} + o_P(1)$ , uniformly in  $t \in T$ .*
- (ii) *Under  $H_0$ ,  $P(S_n > \hat{c}_{1-\alpha}) \rightarrow \alpha$ .*
- (iii) *Under the alternative  $H_1$ ,  $P(S_n > \hat{c}_{1-\alpha}) \rightarrow 1$ .*

The influence-function representation in the bootstrap context is a “re-weighted version” of the influence-function representation obtained in Proposition 1. The new weights are represented by the bootstrap variables  $\{\tilde{\zeta}_i\}_{i=1}^n$ . This ensures that, conditionally on the sample data, the distribution of  $S_n^*$  mimics the behavior of  $S_n$  under  $H_0$ . Conversely, under  $H_1$  the difference  $S_n - \hat{c}_{1-\alpha}$  diverges towards



infinity, ensuring the consistency of the test. In practice, the computation of the critical value  $\hat{c}_{1-\alpha}$  can be performed according to the Monte Carlo procedure outlined at the end of Section 2.

## 6 Applications of the Test Beyond the Benchmark Framework

**Separable single-index model with endogeneity.** Consider the semiparametric model analyzed in [Newey et al. \(1999\)](#). Let  $Y$  be a continuous variable and  $X = (X^e, Z_1)^T$  be a vector of regressors. The structural model takes the form

$$Y = G_0(\beta_0^T X) + \varepsilon, \text{ with} \quad (15)$$

$$E\{\varepsilon | X, Z\} = E\{\varepsilon | V\}.$$

$\varepsilon$  is an unobserved error term,  $G_0(\cdot)$  is a nonparametric structural function,  $\beta_0$  is a parameter of interest for the researcher, and the regressor  $X^e$  is endogenous. The control function  $V$  is defined in the same way as in Eq. (2). The second part of Eq. (15) represents a mean-independence condition<sup>9</sup>. Eq. (15) implies a restriction on the distribution of the data that writes as in Eq. (3), and Remark 2.1 also applies to this context.

**Models with sample selection.** [Escanciano et al. \(2016\)](#) extends the sample-selection model originally proposed in [Heckman \(1979\)](#) to a semiparametric context<sup>10</sup>. Let  $\tilde{Y}$  be a scalar random variable denoting an agent's decision. Assume that such a decision depends on a vector of covariates  $X$  and an unobserved error  $\varepsilon$ . In the presence of sample selection, the agent's decision  $\tilde{Y}$  will be observed only in the selected sample. Let  $D$  denote the selection variable. The model writes as

$$\tilde{Y} = \phi_0(X^T \beta_0, \varepsilon), \quad D = 1\{m_0(Z) \geq u\}. \quad (16)$$

$\phi_0$  is a function known by the researcher, while  $m_0$  is an unknown function.  $u \sim U[0, 1]$ , without loss of generality, so  $m_0$  is identified as  $m_0(Z) = E\{D | Z\}$ . If  $D = 1$  the individual is selected and his decision  $\tilde{Y}$  is observed; if  $D = 0$  the individual is not selected and his decision  $\tilde{Y}$  is not observed. Accordingly, the researcher observes  $Y := \tilde{Y} \cdot D$ .

As an example, it is possible to set  $\phi_0(X^T \beta_0, \varepsilon) = 1\{X^T \beta_0 \geq \varepsilon\}$  or  $\phi_0(X^T \beta_0, \varepsilon) = \max\{0, X^T \beta_0 + \varepsilon\}$ . In the former case, we would have  $\tilde{Y} \in \{0, 1\}$ , and hence a binary-choice model with sample selection. In the latter case,  $\tilde{Y} = \max\{0, X^T \beta_0 + \varepsilon\}$  and we would obtain a truncated regression model where  $\tilde{Y}$  is observed only when taking positive values in the selected sample. This truncated regression model with sample selection is also called “double hurdle-model” ([Escanciano et al., 2016](#); [Cragg, 1971](#)). In the sample-selection literature, the errors  $(\varepsilon, u)$  are assumed to be jointly independent from the set

<sup>9</sup>The identification of both the parameter  $\beta_0$  and the function  $m_0$  is discussed in [Newey et al. \(1999\)](#).

<sup>10</sup>Beyond [Escanciano et al. \(2016\)](#), identification of these models can also be obtained by the results in [Rothe \(2009\)](#) and [Blundell & Powell \(2004\)](#).

of covariates  $(X, Z)$ , i.e.

$$(\varepsilon, u) \perp (X, Z), \quad (17)$$

but they are allowed to be mutually dependent from each other<sup>11</sup>.

The model so far described is a semiparametric version of Heckman's sample-selection model where, following [Escanciano et al. \(2016\)](#), the distribution of  $(\varepsilon, u)$  is left unspecified. Let us denote it with  $G_0$ . Eq. (16) and (17) imply the following restriction on the distribution of the data

$$E\{Y|X, Z\} = G_0(X^T \cdot \beta_0, m_0(Z)). \quad (18)$$

Testing the above equation is equivalent to checking the correct specification of the sample-selection model, and the procedure described in Section 2 can be easily adapted to the present context. Remark 2.1 also applies to this model.

**Binary-choice models with endogeneity and structure on the error term.** A further application of the test consists in the specification of a binary-choice model with two or more endogenous variables. In this case, the test can be applied to check the validity of the restrictions imposed on the error term to reduce the curse of dimensionality. Consider the model

$$Y = 1\{X^T \theta \geq u\}, X := (X_1^e, X_2^e, Z)^T, \quad (19)$$

where  $X_1^e$  and  $X_2^e$  are two endogenous regressors correlated with the unobserved error term  $u$ . In the presence of two endogenous regressors, the researcher needs two control functions to handle endogeneity. So, similarly to Section 2, the control functions are defined as  $V_1 := X_1^e - m_1(Z_1)$  and  $V_2 := X_2^e - m_2(Z_2)$  with  $E\{V_1|Z_1\} = 0$  and  $E\{V_2|Z_2\} = 0$ .  $Z$  is included in  $Z_1$  and  $Z_2$ . Define  $Z := (Z_0, Z_1, Z_2)^T$  and denote with  $(X, Z)$  the vector gathering the components of  $X$  and  $Z$  without repeating the common ones. By imposing  $u|X, Z \sim u|V_1, V_2$ , we obtain  $E\{Y|X, Z\} = E\{Y|X \cdot \theta_0, V_1, V_2\}$ . Thus, the estimation of the vector  $\theta_0$  and the Average Structural Function will require the nonparametric estimation of a function with three arguments,  $E\{Y|X_1 \cdot \theta_0 = \cdot, V_1 = \cdot, V_2 = \cdot\}$ . This is a triple-index model. To reduce the curse of dimensionality and increase the tractability of the framework, it would be useful to impose some structure on the unobserved error term  $u$ . So, assume that

$$u = \gamma_0 V_1 + g_0(V_2) + \varepsilon, \varepsilon \perp (X, Z), \quad (20)$$

where  $g_0$  is an unknown function. Eq. (20) and Eq. (19) imply

$$E\{Y|X, Z\} = G_0(\beta_{0,1}^T X + \gamma_0 m_1(Z_1), X_2^e - m_2(Z_2)), \quad (21)$$

<sup>11</sup>This implies "selection on unobservables," in the sense that even by keeping fixing  $(X, Z)$  the "potential outcome"  $\tilde{Y}$  will still be correlated with the selection variable  $D$  ([Heckman, 1979](#); [Escanciano et al., 2016](#)).

where  $\beta_{0,1} := (\theta_{0,1} - \gamma_0, \theta_{0,2}^T)^T$ ,  $\theta_{0,1}$  is the first component of  $\theta_0$ , while the vector  $\theta_{0,2}$  gathers all the remaining components of  $\theta_0$ . Notice that Eq. (20) implies  $u|X, Z \sim u|V_1, V_2$ . When Eq. (21) holds, the estimation of  $(\beta_{0,1}^T, \gamma_0)$  requires an estimate of the conditional expectation  $E\{Y|X^T \cdot \beta_{0,1} + \gamma_0 m_1(Z_1), X \perp m_2(Z_2)\}$ , which involves two indices. Accordingly, the constraint in Eq. (20) can attenuate the curse of dimensionality and simplify the estimation of the structural elements. The method presented in Section 2 can be easily adapted to the present context<sup>12</sup>.

**Semiparametric games with incomplete information.** [Aradillas-Lopez \(2012\)](#) and [Lewbel & Tang \(2015\)](#) study identification and estimation of a binary-choice game with incomplete information. For simplicity, we assume a context with only two players. Denote each player by the index  $p \in \{1, 2\}$ . Each of them must take a binary decision  $a_p \in \{0, 1\}$ . Let  $X_p$  be the exogenous covariates entering agent  $p$ 's payoff, and assume that these covariates can be observed by each player and the researcher. For notational simplicity, we write  $X = (X, X^T)$ . Each player has private information denoted by  $u_p$  that neither the other agent nor the researcher can observe. It is assumed that  $u_1 \perp u_2$  and  $(u_1, u_2) \perp X$ . Each payer knows the distribution of the other player's private information, while the researcher doesn't. The payoff function of player  $p$  takes the form

$$\Pi_p(a_p, \tilde{X}, a_{-p}, u_p) = a_p \cdot [x_p^T \cdot \gamma_p - \alpha_p \cdot g_{-p}(\tilde{X}) - u_p],$$

where  $\gamma_{0,p}$  is a vector. Assuming that a unique Bayesian-Nash equilibrium is played, the model implies the following semiparametric restriction on the distribution of the data ([Aradillas-Lopez, 2012](#); [Lewbel & Tang, 2015](#)):

$$E\{a_p|\tilde{X}\} = E\{a_p|X_p^T \cdot \gamma_p - \alpha_p \cdot g_{-p}(\tilde{X})\} \text{ with } g_{-p}(\tilde{X}) := E\{a_{-p}|\tilde{X}\} \text{ for } p = 1, 2, \quad (22)$$

which is a specific version of the moment condition in Eq. (21)<sup>13</sup>. The procedure presented in Section 2 can be easily adapted to the present context.

## 7 Simulation Study and Empirical Example

In this section, we provide a Monte Carlo experiment to assess the small-sample performance of the test. We focus on the binary-choice model with control functions of Section 2. We generate  $Y$  from a

<sup>12</sup>The validity of the bootstrap procedure is proved in the Appendix by a general model that encompasses all examples of application analyzed in this paper.

<sup>13</sup>The identification of  $(\gamma_{0,p}, \alpha_{0,p})$  is described in [Aradillas-Lopez \(2010\)](#) and [Lewbel & Tang \(2015\)](#). It essentially requires the presence of specific profit shifters and normalization conditions.

model involving two observables  $(X^e, Z_1)$  and an unobserved component  $u$ ,

$$Y = 1 \{ X^e + \delta \cdot Z_1 + a \cdot (1 - \delta) \cdot (Z_1^2 - 1)/2 \geq u \}.$$

We assume that the researcher specifies the model as

$$Y = 1 \{ X^e + \beta_0 \cdot Z_1 \geq u \} \quad (23)$$

and considers  $\beta_0$  as an unknown parameter. So, when  $\delta = 1$  the model postulated by the researcher is correctly specified, while with  $\delta = 0$  the model is misspecified. The coefficient  $a$  measures the degree of misspecification. The unobserved variable  $u$  is defined as  $u = u^* + V$ , with

$$V := X^e - m_0(Z), \quad m_0(Z) = \alpha_1 Z_1 + \alpha_2 Z_2,$$

$\alpha_1 = \alpha_2 = 1/\sqrt{2}$  and  $Z := (Z_1, Z_2)^T$ . The error  $u^*$  is generated from different distributions specified below and is independent from all the other variables. The functional form of  $m_0$  is unknown to the researcher, so  $V$  must be estimated nonparametrically.  $u$  is not fully independent from the covariates  $X$  and is correlated to the regressor  $X^e$  through  $V$ . This variable controls the correlation between the endogenous regressor and the unobserved error, and plays the role of a control function.

We generate  $Z_1$  and  $V$  from two standard normals independent from each other.  $Z_2$  is resampled from an exponential distribution truncated from above at 3 and standardized to have mean 0 and variance 1. To check the robustness of the test with respect to different DGPs, we consider three specifications for the distribution of  $u^*$ :

$$DGP1) u^* \sim N(0, \text{sd} = \sqrt{7});$$

$$DGP2) u^* \sim \frac{\sqrt{7}}{2.5} \cdot \chi^2_{(5)} - 5;$$

$$DGP3) u^* \sim 0.8 \cdot N(-2.5, \text{sd} = \sqrt{3.5}) + 0.2 \cdot N(2.5, \text{sd} = 1).$$

The first DGP delivers a (rescaled) probit model with a distribution of  $u^*$  that is unimodal and symmetric around zero. The second DGP delivers a unimodal distribution for  $u^*$  with positive asymmetry and left skewness. Finally, the third DGP generates  $u^*$  according to a mixture between two Gaussians and delivers a distribution for  $u^*$  that is bimodal and left-skewed. The three DGPs are built to share some common features under both the null and the alternative hypothesis. Across the three DGPs it holds that  $\text{Var}(u) \approx 8$ ,  $\text{Corr}(u, V) \approx 0.35$ ,  $\text{Corr}(u, X^e) \approx 0.25$ ,  $\text{Var}(X^e + Z_1) \approx 4.5$ .

We assume the researcher wants to test the correct specification of the model he sets up in Eq. (23). Since this equation implies that  $E\{Y | X, Z\} = G_0(X^T \beta_0, V)$ , specification testing can be based on this latter restriction. When  $\delta = 1$  we are under the null hypothesis, while with  $\delta = 0$  we are under  $H_1$ .

## 7.1 Implementation of the Test

The test is implemented as described in Section 2 and Section 3. To get  $\tilde{m}_0$ , we need to set up a bandwidth rate  $h_0$  and a kernel function  $K_0$ . Since  $Z = (Z_1, Z_2)$  has two components, we introduce two bandwidths for the estimation of  $m_0$ : one bandwidth for  $Z_1$  and another for  $Z_2$ . We use a Rule-of-Thumb that is consistent with the rates reported in Section 3, setting  $h_0 = (\hat{s}d(Z_1), \hat{s}d(Z_2)) \cdot n^{-1/6}$ . The kernel  $K_0$  is chosen to be a second-order Gaussian kernel. For the estimator  $\hat{G}_{(\hat{\beta}, \tilde{V})}^{(-i)}$ , we set  $K$  as a second-order Gaussian kernel. Following [Rothe \(2009\)](#), [Delecroix et al. \(2006\)](#), [Escanciano et al. \(2016\)](#), [Maistre & Patilea \(2014\)](#), we compute  $\hat{\beta}$  from the following program

$$(\hat{\beta}, \hat{h}) = \underset{\beta, h}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{G}_{(\hat{\beta}, \tilde{V})}^{(-i)}(X_i^T \beta, \tilde{V}_i)] \cdot \hat{\tau}_i,$$

where  $\hat{G}_{(\hat{\beta}, \tilde{V})}^{(-i)}$  denotes the leave-one-out estimator of  $G_{\beta}$ . We consider the first component of  $(\hat{\beta}, \hat{h})$  as an estimator for  $\beta_0$ . Such a procedure is common in the empirical implementations of semiparametric index models, see [Rothe \(2009\)](#) and [Escanciano et al. \(2016\)](#). To control for the random denominator present in  $\tilde{m}_0$  and  $\hat{G}_{(\hat{\beta}, \tilde{V})}^{(-i)}$ , we introduce the trimming  $\hat{\tau}$  in the above objective function. Specifically,  $\hat{\tau}$  drops those indices  $i$  for which  $|\tilde{V}_i|$  exceeds its 95th quantile. For the numerical minimization, we use a Newton-Ramphson method and select as a starting point for  $\hat{\beta}$  the estimate from a probit model. This model holds true for DGP1, but does not hold for DGP2 or DGP3. The starting point for  $h$  is instead set by a rule of thumb to  $(\hat{s}d(X^T \hat{\beta}_{\text{probit}}), \hat{s}d(\tilde{V})) \cdot n^{-1/8}$ . Although this estimation procedure does not fully respect the restrictions on the bandwidth rates described in Section 3, it is a standard practice in the statistical and econometric literature (see e.g. [Rothe, 2009](#); [Delecroix et al., 2006](#); [Escanciano et al., 2016](#); [Maistre & Patilea, 2014](#); [Xia et al., 2004](#)). Moreover, such an estimation method, when combined with the test, shows a good small-sample performance.

To compute the estimators  $\tilde{G}_{(\hat{\beta}, \tilde{V})}$  and  $\hat{f}$  that will be used in the statistic  $S_n$ , we choose  $K$  as a second-order Gaussian kernel. Since  $(\hat{\beta}^T X, \tilde{V})$  has two components, we introduce two bandwidth rates: one bandwidth for  $\hat{\beta}^T X$  and another bandwidth for  $\tilde{V}$ .  $h$  is set according to a Rule-of-Thumb, so  $h = (\hat{s}d(\hat{\beta}^T X), \hat{s}d(\tilde{V})) \cdot n^{-1/6}$ . Once the estimates  $\tilde{G}_{(\hat{\beta}, \tilde{V})}$  and  $\hat{f}$  are obtained, the computation of the test statistic needs the selection of the weighting function  $\phi$  and the measure  $\mu$ . To obtain a simple expression for  $S_n$ , we set  $\phi(\cdot) = \exp(i \cdot)$  and  $\mu$  to the standard Gaussian. This delivers the following expression for the test statistic:

$$S_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \tilde{\varepsilon}_i \cdot \hat{f}_i \cdot \tilde{\varepsilon}_j \cdot \hat{f}_j \phi(\hat{\beta}^T (X_i - X_j), \tilde{V}_i - \tilde{V}_j), \quad (24)$$

where  $\tilde{\varepsilon} := Y - \tilde{G}_{(\hat{\beta}, \tilde{V})}(\hat{\beta}^T X, \tilde{V})$ ,  $\phi_{\mu}$  is the characteristic function of the standard multivariate Gaussian with dimension  $\dim(X, Z)$ . Since in the literature on specification testing it is common practice to implement Stute's types of tests (see e.g. [Delgado & Manteiga, 2001](#); [Xia et al., 2004](#)), we also evaluate

the performance of a test statistic where  $\varphi_t(x, z) = 1\{(x, z) \leq t\}$  and the measure  $\mu$  is set to the empirical measure  $P_n$ . Such a statistic takes the form

$$S_n^{(St)} = \frac{1}{n^2} \sum_{s=1}^n \sum_{i=1}^n \sum_{j=1}^n \tilde{\varepsilon}_i \cdot \hat{f}_i \cdot 1\{(X, Z)_i \leq (X, Z)_s\} \cdot \tilde{\varepsilon}_j \cdot \hat{f}_j \cdot 1\{(X, Z)_j \leq (X, Z)_s\}. \quad (25)$$

The proofs provided in the Appendix also hold for the test based on the above statistic.

For the bootstrap implementation, we follow the procedure outlined in Section 2 and detailed in Section 5. We set the bootstrap weights  $\{\tilde{\varepsilon}_i\}_{i=1}^n$  to be iidN(0, 1) and compute the critical value  $\hat{c}_{1-\alpha}$  according to the Monte Carlo procedure described at the end of Section 2<sup>14</sup>.

## 7.2 Simulation Results

To evaluate the small sample behavior of the test, we compare it to other procedures aiming at testing a similar null hypothesis. Due to the novelty of the methodology in this paper, no procedure available in the statistical literature can be employed as an alternative to this test in a *real-data* application. Therefore, we choose to compare our methodology to several “Oracle” specification tests. We consider the specification test by [Delgado & Manteiga \(2001\)](#) (DGM now henceforth). The DGM test is designed to check the significance of covariates in a nonparametric regression, assuming the full observability of the regressors. It essentially requires that the model specified under  $H_0$  must be nested into the class of models considered under  $H_1$ . Such a condition holds in our framework<sup>15</sup>.

We here employ the DGM test assuming that the variables  $(X^e + Z_1, V)$  are observed, but the link function  $G_0$  -i.e. the conditional expectation  $E\{Y \mid Z_1 + X^e = \cdot, V = \cdot\}$  is unknown. Such a link function has to be estimated nonparametrically. For the implementation of the DGM test, we follow the suggestions in [Delgado & Manteiga \(2001\)](#). We regress nonparametrically  $Y$  onto  $(Z_1 + X^e, V)$  by kernel smoothing. A bias reducing kernel is employed, so the kernel is chosen to be a 4th order Gaussian kernel. The bandwidth rate is defined as  $(\hat{s}d(Z_1 + X^e), \hat{s}d(V)) \cdot n^{-1/6}$ . We consider two variants of the DGM test. In a first version, we compute the DGM statistic using a Bierens’ characterization of the null hypothesis. This gives rise to a statistic similar to Eq. (24). In a second version, we consider the Stutes’ version of the DGM test that delivers a statistic similar to Eq. (25). The difference between the test presented in this paper and the DGM test is the presence of the nonparametrically generated regressors. The comparison of these two methodologies is useful for understanding the impact of the estimated covariates on the size and power of the test.

In the simulation study, the number of Monte-Carlo replications is set to 1000, the number of bootstrap replications to 100, and for each test we consider the sample sizes  $n = 125, 250, 500$ . The results of the simulations are reported in Tables 1, 2, and 3. They contain the rejection frequencies for each test.

<sup>14</sup>We have also run simulations for the case where the bootstrap weights are set according to the same distribution as in [Delgado & Manteiga \(2001\)](#), but the results essentially do not change.

<sup>15</sup>More precisely, the sigma field generated by the significant covariates under  $H_0$  must be included into the sigma field generated by the covariates under  $H_1$ .

The acronym *NPGR-B* (Nonparametrically Generated Regressors - Bierens) denotes the test proposed in this paper when applied according to Eq. (24). The acronym *NPGR-S* (Nonparametrically Generated Regressors - Stute) denotes the test statistic computed as in Eq. (25). *DGM-S* stands for DGM's test when applied according to the Stute's function, similarly to Eq. (25). Finally, the acronym *DGM-B* stands for DGM's test when applied according to the Bierens' function, similarly to Eq. (24).

On its own, the test presented in this paper performs reasonably well in terms of size and power, across the DGPs considered. When the null hypothesis holds, the test shows a contained error in the rejection probability that becomes smaller as the sample size increases. As regards the power, the test is able to detect departures from the null hypothesis with good frequencies. When the sample size increases, the probability of rejecting the null grows.

When compared to the Oracle DGM Test, the test performs well in terms of size but it is outperformed in terms of power. Specifically, under the null, both the test we propose and the DGM test need an increase in sample size to reduce the error in the rejection probability. The Oracle DGM test shows a better capacity for detecting departures from the null hypothesis compared to the *NPGR-B* and *NPGR-S* tests for any sample size. Also, as the sample size grows, the increase in the probability of rejecting the null seems to be more pronounced for the Oracle DGM test than for the *NPGR-B* and *NPGR-S* tests.

Within this simulation experiment, the presence of nonparametrically generated regressors does not seem to have a considerable impact on the empirical size of the test. Instead, it has an impact on its power and reduces its capacity to detect departures from the null hypothesis, although the test still shows a reasonable ability in detecting the alternative.

### 7.3 An Empirical Application

As an empirical application we test the specification of a model for married women's labor force participation, see [Wooldridge \(2015\)](#). Such a model relates the decision on labor force participation to "other sources of income" of the household. The initial data set consists of 5634 observations coming from the 1991 Current Population Survey<sup>16</sup>. We select the observation on the basis of the experience, including only those women whose level of experience is lower than the 75th percent quantile and larger than the 25th percent quantile of the total distribution. After this change, the sample consists of 2762 observations.

The dependent variable  $Y$  is an indicator that equals one if the woman participates in the labor force and zero if not. This decision depends on "other sources of income" of the household ( $X_{nwifcinc}^e$ ). Other controls in  $X$  are the woman's level of experience ( $Z_{exper}$ ), its square ( $Z_{exper}^2$ ), education ( $Z_{educ}$ ), and a dummy variable that equals one if a child under the age of six is part of the household ( $Z_{kidslt6}$ ). Following [Wooldridge \(2015\)](#), we treat  $X_{nwifcinc}^e$  as endogenous, since unobserved elements might directly affect labor force participation ( $Y$ ) and be correlated with  $X_{nwifcinc}^e$ . To handle the endogeneity, [Wooldridge \(2015\)](#) makes use of the husband's level of education ( $Z_{huseduc}$ ). This variable is as-

---

<sup>16</sup>I am very grateful to Jeffrey Wooldridge for having shared his data set.



sumed to impact the women's labor force participation only indirectly, through the endogenous variable  $X_{nwifcinc}^e$ . Thus, the exclusion restriction necessary for the identification is satisfied. A linear regression of  $X_{nwifcinc}^e$  onto the exogenous variables  $Z := (Z_{exper}, Z_{educ}, Z_{kidl6}, Z_{huseduc})$  shows that the coefficient attached to the exogenous instrument  $Z_{huseduc}$  is significant (see Table 5).

We normalize to unity the coefficient attached to experience,

$$Y = 1\{Z_{exper} + \beta_{0,1}Z_{exper}^2 + \beta_{0,2}Z_{educ} + \beta_{0,4}Z_{kidl6} + \beta_{0,3}X_{nwifcinc}^e \geq u\}$$

$$X_{nwifcinc}^e = m_0(Z) + V, E\{V|Z\} = 0.$$

Following a semiparametric approach, we do not specify the distribution of  $(u, V)$ . Also,  $m_0$  is not restricted to have a specific functional form and is nonparametrically specified. We adopt a normalization on the continuous variable  $Z_{exper}$  that we expect to positively impact the probability of labor force participation. We also expect  $Z_{educ}$  to have a positive impact on such probability. Conversely, we expect both  $Z_{kidl6}$  and  $X_{nwifcinc}^e$  to have a negative impact on the labor force participation decision.

To check the correct specification of the model, we test the null hypothesis introduced in Section 2. The components of the test statistic are obtained as follows. First, we estimate the regression  $m_0$  nonparametrically. We make use of the `np` package by Hayfield & Racine (2008) in R. We select the bandwidths by cross-validation and employ a second-order Gaussian kernel. In the nonparametric first-step estimation, only experience is treated as a continuous regressor, while the remaining ones are treated as discrete variables. Due to the introduction of the bias corrections for the nonparametric estimators, the cross validation method for bandwidths selection is fully coherent with our testing procedure.

Once the residuals  $\tilde{V}$  are computed, we estimate  $\beta_0$  by minimizing a Semiparametric Least-Square criterion as in Section 3. We carry on a numerical optimization using the `BFGS` method in R. The estimates obtained are reported in the second part of Table 5. The sign of each estimated coefficient is as we expected<sup>17</sup>.

In a third step, the function  $G_0$  and the density  $f$  are constructed as presented in the previous sections. According to the guidelines provided in the simulations, we set the bandwidths according to a rule of thumb and employ second-order Gaussian kernels. Similarly to Section 7, we implement two types of tests: a Bierens' test and a Stute's test. The bootstrap procedure is carried over by resampling the bootstrap weights  $\{\xi_i\}_{i=1}^n$  from a standard normal distribution. At each bootstrap iteration, we follow the same steps as those performed at the sample level. However, as described in Section 5, in the bootstrap context the bias corrections do not need to be estimated at each bootstrap iterations are computed with

<sup>17</sup>The benchmark initial values used to start the estimation are the coefficients obtained from an endogenous parametric probit, where the endogeneity is handled by a control function estimated in a first step by a linear regression. The initial values for the bandwidths are set to the values obtained from the estimation of a single-index semiparametric model, where  $\tilde{V}$  is considered as a regressor and both the coefficients and the bandwidths are obtained by minimizing a SLS criterion. This is carried out by the `npindex` function in the `np` package in R. Given the benchmark initial values, in order to avoid the convergence to local minima, we run the minimization several times by considering as initial starting points half, twice, three, and four times the benchmark initial values. We finally select the estimate of  $\beta_0$  delivering the minimum value for the objective function.



sample data. To compare our results, we also implement a specification test for the endogenous probit model where the control function is estimated in a preliminary step by a linear regression. The benchmark results are reported in Table 6. As a robustness check, we report in Table 7 the results of a test where the benchmark bandwidth is multiplied by a scaling factor  $C$ .

The correct specification of the *fully* parametric probit is rejected. Differently, with the procedure we propose, the correct specification of the semiparametric model cannot be rejected at the 2, 3, or 4 percent critical level. Considering the large size of the sample, such results are likely not affected by a small-sample problem generating a lack of power of the test. The different results obtained by our test and the parametric test are not surprising. The parametric probit model imposes stringent parametric constraints that are not fully justified from a theoretical standpoint. The semiparametric framework relaxes some of these constraints, although it imposes a linearity restriction on the index that is not economically justified.

## 8 Conclusions

This paper presents a novel wild bootstrap specification test for semiparametric models with nonparametrically generated regressors. The range of application of the test is wide and includes semiparametric models with endogenous regressors identified by control functions, semiparametric binary-choice models with a nonparametric selection mechanism, truncated variable models with sample selection, and semiparametric games with incomplete information.

We prove the validity of the test under low-level conditions. The statistic contains a bias correction for the nonparametric estimators that allows to implement the test without undersmoothing. Such a bias correction is novel for semiparametric models with generated variables and is attractive because it widens the spectrum of kernel orders and bandwidth rates admissible for the test. We have proposed a Silverman’s Rule of Thumb for the bandwidth selection, and a set of Monte-Carlo simulations has shown that the test performs reasonably well in small samples both in terms of size and power.

One drawback of the bootstrap test of this paper is that it might be computationally demanding. The finite-dimensional parameter must be estimated by a Semiparametric Least-Square method. This implies that at every bootstrap iteration a nonlinear function must be minimized. Consequently the application of the bootstrap test might involve a relevant amount of time. This drawback, however, is not specific to the methodology of this paper, but concerns all those tests for nonlinear models where the finite-dimensional parameter is estimated by a nonlinear optimization.

As a byproduct, we obtain that the Wild Bootstrap scheme proposed in this paper provides a valid inference for the finite-dimensional parameter in a class of semiparametric models with generated regressors. Bootstrap methods are relevant in these frameworks, as the asymptotic covariance matrix of such models contains nonparametric derivatives. Estimating such asymptotic covariances might yield poor confidence intervals because the nonparametric estimators of the derivatives generally have slow rates of convergence. Instead, the Wild Bootstrap procedure proposed in this paper can be considered

as a valuable alternative to this asymptotic approximation. It can also be considered as a valuable alternative to the pairwise bootstrap which does not exploit the restriction the model imposes on the distribution of the data. The wild-bootstrap scheme, instead, uses the information suggested by the setup. Exploring the performance of this method is an interesting topic for future research.

## References

- Ahn, Hyungtaik. 1997. Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors. *Econometric theory*, **13**(1), 3–31.
- Ahn, Hyungtaik, & Manski, Charles F. 1993. Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of econometrics*, **56**(3), 291–321.
- Ai, Chunrong, & Chen, Xiaohong. 2003. Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions. *Econometrica*, **71**(6), 1795–1843.
- Aradillas-Lopez, Andres. 2010. Semiparametric estimation of a simultaneous game with incomplete information. *Journal of econometrics*, **157**(2), 409–431.
- Aradillas-Lopez, Andres. 2012. Pairwise-difference estimation of incomplete information games. *Journal of econometrics*, **168**(1), 120–140.
- Aradillas-Lopez, Andres, Honore, Bo E., & Powell, James L. 2007. Pairwise Difference Estimation with Nonparametric Control Variables. *International economic review*, **48**(4), 1119–1158.
- Bierens, Herman J. 1982. Consistent model specification tests. *Journal of econometrics*, **20**(1), 105–134.
- Bierens, Herman J., & Ploberger, Werner. 1997. Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica*, **65**(5), 1129–1152.
- Blundell, Richard, & Matzkin, Rosa L. 2014. Control functions in nonseparable simultaneous equations models. *Quantitative economics*, **5**(2), 271–295.
- Blundell, Richard W., & Powell, James L. 2004. Endogeneity in Semiparametric Binary Response Models. *Review of economic studies*, **71**(3), 655–679.
- Bravo, Francesco, Escanciano, Juan Carlos, & Keilegom, Ingrid. 2018. Two-step semiparametric empirical likelihood inference. *Annals of statistics*, July.
- Chen, Xiaohong, Linton, Oliver, & Keilegom, Ingrid Van. 2003. Estimation of Semiparametric Models when the Criterion Function Is Not Smooth. *Econometrica*, **71**(5), 1591–1608.
- Cragg, John G. 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, **39**(5), 829–844.

- Davidson, Russell, & MacKinnon, James G. 2010. Wild Bootstrap Tests for IV Regression. *Journal of business & economic statistics*, **28**(1), 128–144.
- Delecroix, Michel, Hristache, Marian, & Patilea, Valentin. 2006. On semiparametric M-estimation in single-index regression. *Journal of statistical planning and inference*, **136**(Mar.), 730–769.
- Delgado, Miguel, Dominguez, Manuel, & Lavergne, Pascal. 2006. Consistent Tests of Conditional Moment Restrictions. *Annales d'économie et de statistique*, **81**(Jan.), 33–67.
- Delgado, Miguel A., & Manteiga, Wenceslao Gonzalez. 2001. Significance testing in nonparametric regression based on the bootstrap. *The annals of statistics*, **29**(5), 1469–1507.
- Delgado, Miguel A., & Stute, Winfried. 2008. Distribution-free specification tests of conditional models. *Journal of econometrics*, **143**(1), 37–55.
- Einmahl, John H. J., & Van Keilegom, Ingrid. 2008. Specification tests in nonparametric regression. *Journal of econometrics*, **143**(1), 88–102.
- Escanciano, Juan Carlos, & Song, Kyungchul. 2010. Testing single-index restrictions with a focus on average derivatives. *Journal of econometrics*, **156**(2), 377–391.
- Escanciano, Juan Carlos, Jacho-Chávez, David T., & Lewbel, Arthur. 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of econometrics*, **178**(P3), 426–443.
- Escanciano, Juan Carlos, Jacho Chavez, David, & Lewbel, Arthur. 2016. Identification and estimation of semiparametric two-step models. *Quantitative economics*, **7**(2), 561–589.
- Escanciano, Juan Carlos, Pardo-Fernandez, Juan Carlos, & Keilegom, Ingrid Van. 2018. Asymptotic distribution-free tests for semiparametric regressions with dependent data. *The annals of statistics*, **46**(3), 1167–1196.
- Fan, Yanqin, & Li, Qi. 1996. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica*, **64**(4), 865–890.
- Hahn, Jinyong, & Ridder, Geert. 2013. Asymptotic Variance of Semiparametric Estimators With Generated Regressors. *Econometrica*, **81**(1), 315–340.
- Hahn, Jinyong, Liao, Zhipeng, & Ridder, Geert. 2018. Nonparametric Two-Step Sieve M Estimation and Inference. *Econometric theory*, **34**(6), 1281–1324.
- Hayfield, Tristen, & Racine, Jeffrey S. 2008. Nonparametric Econometrics: The np Package. *Journal of statistical software*, **27**(1), 1–32.

- Heckman, James J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, **47**(1), 153–161.
- Horowitz, Joel L. 2006. Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables. *Econometrica*, **74**(2), 521–538.
- Huang, Meng, Sun, Yixiao, & White, Halbert. 2016. A Flexible Nonparametric Test for Conditional Independence. *Econometric theory*, **32**(Dec.), 1434–1482.
- Ichimura, Hidehiko. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of econometrics*, **58**(1), 71–120.
- Ichimura, Hidehiko, & Lee, Sokbae. 2010. Characterization of the asymptotic distribution of semiparametric M-estimators. *Journal of econometrics*, **159**(2), 252–266.
- Imbens, Guido W., & Newey, Whitney K. 2009. Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity. *Econometrica*, **77**(5), 1481–1512.
- Jones, M. C., & Signorini, D. F. 1997. A Comparison of Higher-Order Bias Kernel Density Estimators. *Journal of the american statistical association*, **92**(439), 1063–1073.
- Kosorok, Michael R. 2007. *Introduction to Empirical Processes and Semiparametric Inference*. Springer.
- Lavergne, Pascal, & Patilea, Valentin. 2008. Breaking the curse of dimensionality in nonparametric testing. *Journal of econometrics*, **143**(1), 103–122.
- Lavergne, Pascal, & Patilea, Valentin. 2013. Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory. *Journal of econometrics*, **177**(1), 47–59.
- Lavergne, Pascal, Maistre, Samuel, & Patilea, Valentin. 2015a. A significance test for covariates in nonparametric regression. *Electronic journal of statistics*, **9**(1), 643–678.
- Lavergne, Pascal, Maistre, Samuel, & Patilea, Valentin. 2015b. A significance test for covariates in nonparametric regression. *Electronic journal of statistics*, **9**(1), 643–678.
- Lewbel, Arthur, & Tang, Xun. 2015. Identification and estimation of games with incomplete information using excluded regressors. *Journal of econometrics*, **189**(1), 229–244.
- Li, Qi, & Racine, Jeffrey Scott. 2006. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Li, Qi, & Wooldridge, Jeffrey M. 2002. Semiparametric Estimation Of Partially Linear Models For Dependent Data With Generated Regressors. *Econometric theory*, **18**(03), 625–645.

- Maistre, Samuel, & Patilea, Valentin. 2014. Nonparametric model checks of single-index assumptions. *Statistica sinica*, **29**(Oct.).
- Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2012. Nonparametric regression with nonparametrically generated covariates. *The annals of statistics*, **40**(2), 1132–1170.
- Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2016. Semiparametric Estimation with Generated Covariates. *Econometric theory*, **32**(05), 1140–1177.
- Neumeyer, Natalie, & Van Keilegom, Ingrid. 2010. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of multivariate analysis*, **101**(5), 1067–1078.
- Newey, Whitney K. 1994. The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, **62**(6), 1349–1382.
- Newey, Whitney K., Powell, James L., & Vella, Francis. 1999. Nonparametric Estimation of Triangular Simultaneous Equations Models. *Econometrica*, **67**(3), 565–604.
- Newey, Whitney K., Hsieh, Fushing, & Robins, James M. 2004. Twicing Kernels and a Small Bias Property of Semiparametric Estimators. *Econometrica*, **72**(3), 947–962.
- Pagan, Adrian. 1984. Econometric Issues in the Analysis of Regressions with Generated Regressors. *International economic review*, **25**(1), 221–247.
- Pollard, D. 1984. *Convergence of Stochastic Processes*. Springer.
- Rivers, Douglas, & Vuong, Quang H. 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, **39**(3), 347–366.
- Rothe, Christoph. 2009. Semiparametric estimation of binary response models with endogenous regressors. *Journal of econometrics*, **153**(1), 51–64.
- Sherman, Robert P. 1994. Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators. *The annals of statistics*, **22**(1), 439–459.
- Stinchcombe, Maxwell B., & White, Halbert. 1998. Consistent Specification Testing With Nuisance Parameters Present Only Under The Alternative. *Econometric theory*, **14**(03), 295–325.
- Vaart, A. W. van der. 1998(Oct.). *Asymptotic Statistics by A. W. van der Vaart*.
- van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge University Press.
- van der Vaart, A. W., & Wellner, Jon. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.

- Vanhems, Anne, & Keilegom, Ingrid Van. 2019. Estimation of a Semiparametric Transformation Model in the Presence of Endogeneity. *Econometric theory*, **35**(1), 73–110.
- W. K. Andrews, Donald. 1994. Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity. *Econometrica*, **62**(Feb.), 43–72.
- Wooldridge, Jeffrey M. 2015. Control Function Methods in Applied Econometrics. *Journal of human resources*, **50**(2), 420–445.
- Xia, Yingcun, Li, Wai, Tong, Howell, & Zhang, Dixin. 2004. A goodness-of-fit test for single-index models. (With comments and rejoinder). *Statistica sinica*, **14**(Jan.).

Table 1: Simulation results for DGP1

| $\alpha$ |        | $H_0$ |       | $H_1$ |       |
|----------|--------|-------|-------|-------|-------|
|          |        | 0.05  | 0.1   | 0.05  | 0.1   |
| $n=125$  | NPGR-B | 0.076 | 0.115 | 0.174 | 0.248 |
|          | NPGR-S | 0.069 | 0.098 | 0.148 | 0.223 |
|          | DGM-B  | 0.073 | 0.136 | 0.178 | 0.295 |
|          | DGM-S  | 0.069 | 0.127 | 0.256 | 0.370 |
|          | DGM-S  | 0.061 | 0.126 | 0.349 | 0.625 |
| $n=250$  | NPGR-B | 0.071 | 0.104 | 0.269 | 0.361 |
|          | NPGR-S | 0.062 | 0.091 | 0.223 | 0.326 |
|          | DGM-B  | 0.059 | 0.127 | 0.320 | 0.444 |
|          | DGM-S  | 0.064 | 0.121 | 0.469 | 0.585 |
|          | DGM-S  | 0.054 | 0.094 | 0.330 | 0.433 |
| $n=500$  | NPGR-B | 0.041 | 0.067 | 0.273 | 0.414 |
|          | NPGR-S | 0.061 | 0.125 | 0.570 | 0.693 |
|          | DGM-B  | 0.061 | 0.125 | 0.570 | 0.693 |
|          | DGM-S  | 0.061 | 0.125 | 0.570 | 0.693 |
|          | DGM-S  | 0.061 | 0.125 | 0.570 | 0.693 |

Notes: Simulations based on 1000 Monte-Carlo replications. The test is based on 1000 bootstrap samples.  $h_0 = (sd(Z_1), sd(Z_2)) \cdot n^{-1/6}$ ,  $h = (sd(X^T \cdot \hat{\beta}), sd(\hat{V}))$ . The kernels are Gaussian Kernels of order 2 for the NPGR test, and Gaussian kernels of order 4 for the DGM tests. For  $H_0$ , ( $\delta = 1, a = 0$ ); for  $H_1$ , ( $\delta = 0, a = 3/2$ ).

Table 2: Simulation results for DGP2

| $\alpha$ |        | $H_0$ |       | $H_1$ |       |
|----------|--------|-------|-------|-------|-------|
|          |        | 0.05  | 0.1   | 0.05  | 0.1   |
| $n=125$  | NPGR-B | 0.078 | 0.116 | 0.195 | 0.283 |
|          | NPGR-S | 0.075 | 0.104 | 0.185 | 0.281 |
|          | DGM-B  | 0.076 | 0.153 | 0.213 | 0.316 |
|          | DGM-S  | 0.066 | 0.153 | 0.307 | 0.427 |
|          | DGM-S  | 0.061 | 0.108 | 0.354 | 0.486 |
| $n=250$  | NPGR-B | 0.069 | 0.117 | 0.304 | 0.395 |
|          | NPGR-S | 0.062 | 0.097 | 0.277 | 0.379 |
|          | DGM-B  | 0.05  | 0.119 | 0.370 | 0.513 |
|          | DGM-S  | 0.062 | 0.112 | 0.549 | 0.668 |
|          | DGM-S  | 0.061 | 0.108 | 0.354 | 0.486 |
| $n=500$  | NPGR-B | 0.047 | 0.083 | 0.316 | 0.448 |
|          | NPGR-S | 0.057 | 0.119 | 0.632 | 0.745 |
|          | DGM-B  | 0.061 | 0.108 | 0.354 | 0.486 |
|          | DGM-S  | 0.061 | 0.108 | 0.354 | 0.486 |
|          | DGM-S  | 0.061 | 0.108 | 0.354 | 0.486 |

Notes: Simulations based on 1000 Monte-Carlo replications. The test is based on 1000 bootstrap samples.  $h_0 = (sd(Z_1), sd(Z_2)) \cdot n^{-1/6}$ ,  $h = (sd(X^T \cdot \hat{\beta}), sd(\hat{V}))$ . The kernels are Gaussian Kernels of order 2 for the NPGR test, and Gaussian kernels of order 4 for the DGM test. For  $H_0$ , ( $\delta = 1, a = 0$ ); for  $H_1$ , ( $\delta = 0, a = 3/2$ ).

Table 3: Simulation results for DGP3

| $\alpha$ |        | $H_0$ |       | $H_1$ |       |
|----------|--------|-------|-------|-------|-------|
|          |        | 0.05  | 0.1   | 0.05  | 0.1   |
| $n=125$  | NPGR-B | 0.078 | 0.113 | 0.197 | 0.287 |
|          | NPGR-S | 0.078 | 0.106 | 0.188 | 0.264 |
|          | DGM-B  | 0.068 | 0.132 | 0.193 | 0.307 |
|          | DGM-S  | 0.062 | 0.130 | 0.283 | 0.384 |
|          |        |       |       |       |       |
| $n=250$  | NPGR-B | 0.063 | 0.10  | 0.297 | 0.392 |
|          | NPGR-S | 0.051 | 0.085 | 0.269 | 0.372 |
|          | DGM-B  | 0.064 | 0.119 | 0.348 | 0.475 |
|          | DGM-S  | 0.059 | 0.126 | 0.519 | 0.640 |
|          |        |       |       |       |       |
| $n=500$  | NPGR-B | 0.076 | 0.109 | 0.328 | 0.453 |
|          | NPGR-S | 0.049 | 0.084 | 0.280 | 0.427 |
|          | DGM-B  | 0.054 | 0.107 | 0.603 | 0.710 |
|          | DGM-S  | 0.062 | 0.110 | 0.571 | 0.655 |
|          |        |       |       |       |       |

Notes: Simulations based on 1000 Monte-Carlo replications. The test is based on 100 bootstrap samples.  $h_0 = (sd(Z_1), sd(Z_2)) \cdot n^{-1/6}$ ,  $h = (sd(X^T \cdot \hat{\beta}), sd(\hat{V})) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2 for the NPGR test, and Gaussian kernels of order 4 for the DGM test. For  $H_0$ , ( $\delta = 1, a = 0$ ); for  $H_1$ , ( $\delta = 0, a = 3/2$ ).



Table 4: Simulation results: comparison between the test with bias correction and without bias correction

| <i>C</i>      |      | <i>NPGR B</i> <i>NPGR-B wBC</i> |            |             |            | <i>NPGR S</i> <i>NPGR-S wBC</i> |            |             |            |
|---------------|------|---------------------------------|------------|-------------|------------|---------------------------------|------------|-------------|------------|
|               |      | <b>0.05</b>                     | <b>0.1</b> | <b>0.05</b> | <b>0.1</b> | <b>0.05</b>                     | <b>0.1</b> | <b>0.05</b> | <b>0.1</b> |
| <i>n</i> =125 | 0.5  | 0.035                           | 0.066      | 0.045       | 0.75       | 0.027                           | 0.049      | 0.033       | 0.055      |
|               | 0.75 | 0.043                           | 0.074      | 0.053       | 0.082      | 0.04                            | 0.058      | 0.049       | 0.065      |
|               | 1    | 0.061                           | 0.094      | 0.073       | 0.105      | 0.053                           | 0.074      | 0.068       | 0.096      |
|               | 1.25 | 0.073                           | 0.11       | 0.098       | 0.15       | 0.072                           | 0.098      | 0.097       | 0.138      |
| <i>n</i> =250 | 0.5  | 0.043                           | 0.081      | 0.03        | 0.071      | 0.032                           | 0.064      | 0.022       | 0.038      |
|               | 0.75 | 0.047                           | 0.081      | 0.033       | 0.061      | 0.037                           | 0.057      | 0.025       | 0.034      |
|               | 1    | 0.052                           | 0.09       | 0.044       | 0.064      | 0.046                           | 0.064      | 0.033       | 0.045      |
|               | 1.25 | 0.068                           | 0.113      | 0.062       | 0.093      | 0.062                           | 0.089      | 0.046       | 0.069      |

Notes: Simulations based on 1000 Monte-Carlo replications. The test is based on 100 bootstrap samples.  $h_0 = (sd(Z_1), sd(Z_2)) \cdot n^{-1/6}$ ,  $h = (sd(X^T \cdot \hat{\beta}), sd(\hat{V})) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

Table 5: Descriptive statistics and estimation results

|                       | <i>inlf</i> | <i>exper</i> | <i>exper</i> <sup>2</sup> | <i>nwifeinc</i> | <i>educ</i> | <i>kidlt6</i> | <i>huseduc</i>     |
|-----------------------|-------------|--------------|---------------------------|-----------------|-------------|---------------|--------------------|
| <i>Mean</i>           | 0.612       | 19.82        | 412                       | 32.38           | 13.2        | 0.24          | 13.44              |
| <i>Std. Dev.</i>      | 0.49        | 4.45         | 180                       | 28.23           | 2.46        | 0.42          | 2.91               |
| <i>Min</i>            | 0           | 13           | 169                       | 0               | 0           | 0             | 0                  |
| <i>Max</i>            | 1           | 28           | 784                       | 112.5           | 18          | 1             | 18                 |
| <i>1st stage reg.</i> | -           | 0.002        | -                         | -               | 0.023       | 0.83          | <<10 <sup>-5</sup> |
| <i>SLS est.</i>       | -           | 1            | -0.93                     | -0.23           | 1.19        | -1.74         | -                  |

Notes: The variable *inlf* denotes *Y*. The row *1st stage reg.* contains the results obtained from the linear regression of *nwifeinc* onto *Z* =(*exper*, *educ*, *kidlt6*, *huseduc*). The row *SLS est.* contains the estimation results for the Semiparametric Least Squares method.

Table 6: Main empirical results

|                         | <i>Statistic</i> | <i>Q 90</i> | <i>Q 95</i> | <i>P value</i> |
|-------------------------|------------------|-------------|-------------|----------------|
| <b><i>NPGR-B</i></b>    | 0.4866           | 0.2499      | 0.469       | 0.05           |
| <b><i>NPGR-S</i></b>    | 1.5275           | 1.1601      | 1.9647      | 0.06           |
| <b><i>Bierens-P</i></b> | 0.301            | 0.1935      | 0.2087      | 0              |
| <b><i>Stute-P</i></b>   | 0.0235           | 0.0156      | 0.0186      | 0.021          |

Notes: *Bierens-P* denotes Bierens' test for a parametric model with a linear control function, when applied according to the complex exponential. *Stute-P* denotes Stute's test for the parametric probit with a linear control function. *Q 90* and *Q 95* denote the 90th and the 95th quantile, respectively, of the bootstrap distribution. *Pvalue* denotes the bootstrap p-value.

Table 7: Empirical results: robustness checks

| <i>C</i>           | <i>Test</i>          | <i>Statistic</i> | <i>Q 90</i> | <i>Q 95</i> | <i>P-value</i> |
|--------------------|----------------------|------------------|-------------|-------------|----------------|
| <b><i>0.5</i></b>  | <b><i>NPGR-B</i></b> | 13.2627          | 9.6898      | 14.6649     | 0.06           |
|                    | <b><i>NPGR-S</i></b> | 41.6119          | 43.08432    | 74.9582     | 0.115          |
| <b><i>0.75</i></b> | <b><i>NPGR-B</i></b> | 1.9377           | 1.3402      | 1.7148      | 0.05           |
|                    | <b><i>NPGR-S</i></b> | 6.029            | 4.7391      | 8.3056      | 0.085          |
| <b><i>1.25</i></b> | <b><i>NPGR-B</i></b> | 0.1643           | 0.0876      | 0.1214      | 0.03           |
|                    | <b><i>NPGR-S</i></b> | 0.5344           | 0.3410      | 0.5073      | 0.05           |

Notes: Results from the application of the test with different values of the constant  $C$  multiplying the bandwidth  $h$ .

# Appendix

In this Appendix, we prove our main results. First, we consider a general moment condition that encompasses all the examples of application of our test. The proofs for the asymptotic analysis are contained in Appendix A. Appendix B contains the proofs for the bootstrap test, while Appendix C provides the power analysis. Appendix D contains some auxiliary results that are used throughout the proofs. A **Supplementary Material** that can be found on the author's website contains those technical proofs that are omitted for reasons of space. Let us start by rewriting the null hypothesis as

$$H_0: E\{Y | \tilde{X}\} = E\{Y | \beta_{0,1}^T X_1 + \beta_{0,2}^T m_2(X_2), X^e - m_0(Z)\} \text{ for some } \beta_0 \in \mathbb{R}^{p_1+1},$$

where  $\beta_0 = (\beta_{0,1}^T, \beta_{0,2}^T)^T$ ,  $m_0(Z) := E\{D | Z\}$ ,  $m_2(X_2) := E\{D_2 | X_2\}$ , and  $\tilde{X}$  gathers all the components of  $X_1$ ,  $X_2$ , and  $Z$  without repeating the common ones. I assume for simplicity that  $m_0(Z)$  and  $m_2(X_2)$  are scalar random variables, as the extension to any finite dimension larger than one is straightforward. The variables  $(Y, \tilde{X}, D_0, D_2)$  are observed, while the functions  $m_0$ ,  $m_2$ , and the finite dimensional parameter  $\beta_0$  are unknown. I here consider an omnibus test, so the alternative hypothesis  $H_1$  is defined as the logical complement of  $H_0$ .

**Notation.** Let  $X$  be the vector gathering  $X_1$  and  $X_2$  without repeating the common components.  $X^e$  is a sub-vector of  $X$ . Unless it is differently stated, the capital letters will denote random variables, while the lower cases will denote realizations of random variables. For instance, given the random variable  $X$ , the low case  $x$  will denote a specific realization of  $X$ . The dimension of each random variable is defined as follows:  $p_1 :=$

$\dim(x_1)$ ,  $p_2 := \dim(x_2)$ ,  $p_0 := \dim(z)$ ,  $d := 1 + \dim(v)$ . Let  $\beta_0 := (\beta_1^T, \beta_2^T)^T$  denote the "true" value of the parameters, and  $\beta := (\beta_1^T, \beta_2^T)^T$  denote a generic vector in  $\mathbb{R}^{p_1+1}$ . For any element  $\beta \in \mathbb{R}^{p_1+1}$ , define  $w(\beta) := (\beta_1^T x_1 + \beta_2^T m_2(x_2), x^e - m_0(z))$  and  $W_i(\beta) := (\beta_1^T X_{i,1} + \beta_2^T m_2(X_{i,2}) - X_{i,3} - m_0(Z))$ , with  $i = 1, \dots, n$ , and  $n$  denoting the sample size. Also, let  $f_{W(\beta)}$  denote the density function of the random variable  $W(\beta)$ , and  $tt_{W(\beta)}(w) := E\{Y | W(\beta) = w\}$ . In line with the usual notation of Empirical Process Theory, I define the linear operators  $P$  and  $P_n$  as  $Pg = \int g(y, \tilde{x})P(dy, \tilde{x})$  and  $P_n g = (1/n) \sum_{i=1}^n g(Y_i, \tilde{X}_i)$ , respectively. Notice that if  $(y, \tilde{x}) \mapsto g(y, \tilde{x})$  is a nonrandom and deterministic function,  $Pg = Eg(Y, \tilde{X})$ . For any object  $\gamma$  or  $\gamma_0$ , either functions or vectors,  $\hat{\gamma}$  will denote the estimate of either  $\gamma$  or  $\gamma_0$ . For instance,  $\hat{\beta}$  denotes the estimate of  $\beta_0$ . Also, in line with the notation in Empirical Process Theory, for any set  $A$  containing a countably many elements I denote with  $\#A$  the cardinality of a countable set  $A$ .

Denote with  $\partial tt_{W(\beta)}^T = (\partial_1 tt_{W(\beta)}, \partial_2^T tt_{W(\beta)})$  the partial derivative of  $tt_{W(\beta)}$  with respect to all its  $d$  arguments, so that  $\partial_1 tt_{W(\beta)}$  will denote the partial derivative of  $tt_{W(\beta)}$  with respect to its first argument, while  $\partial_2^T tt_{W(\beta)} = (\partial_2 tt_{W(\beta)}, \dots, \partial_d tt_{W(\beta)})^T$  will denote the vector gathering the partial derivatives of  $tt_{W(\beta)}$  with respect to the other  $d_V$  arguments. Notice that  $\partial tt_{W(\beta)}$  is measurable with respect to (the sigma field generated by)  $W(\beta)$ .

The conditional expectation  $tt_{W(\beta)}(w) := E\{Y | W(\beta) = w\}$  can be seen as a mapping  $(w, \beta) \mapsto tt_{W(\beta)}(w)$ . Notice that a change in  $\beta$  will change the shape of the function  $w \mapsto tt_{W(\beta)}(w)$ . Denote the derivative of the mapping  $(w, \beta) \mapsto tt_{W(\beta)}(w)$  with respect to  $\beta$  as  $\partial_\beta tt_{W(\beta)}(\cdot)$ . Notice that this function is measurable with

respect to (the sigma field generated by)  $W(\beta)$ . On the other hand, for the mapping  $(\tilde{x}, \beta) \mapsto tt_{W(\beta)}(w(\beta))$ , a change in  $\beta$  will impact on both the shape of the function  $tt_{W(\beta)}(\cdot)$  and the argument  $w(\beta)$  where such a function is computed. The differentiation of the mapping  $(\tilde{x}, \beta) \mapsto tt_{W(\beta)}(w(\beta))$  with respect to both occurrences of  $\beta$  will deliver  $\partial_{\beta} tt_{W(\beta)}(w(\beta)) + \partial_1 tt_{W(\beta)}(w(\beta)) \cdot (x_1^T, m_{\tilde{x}})$ . Let us denote such a derivative with  $\nabla_{\beta} tt_{W(\beta)}(w(\beta))$ . Observe that  $\tilde{x} \mapsto \nabla_{\beta} tt_{W(\beta)}(w(\beta))$  is *not* measurable with respect to the sigma field generated by  $W(\beta)$ .

With this new notation, the estimators are now given by

$$\begin{aligned} \hat{m}_0(\tilde{x}) &:= \frac{\hat{T}_0(\tilde{x})}{\hat{f}_0(\tilde{x})}, \quad \hat{T}_0(\tilde{x}) := \frac{1}{nh^{p_0}} \sum_{i=1}^n D_{0,K_0} \cdot \frac{Z_{i-} - \tilde{x}}{h_0}, \quad \hat{f}_0(\tilde{x}) := \frac{1}{nh^{p_0}} \sum_{i=1}^n K_0 \cdot \frac{Z_{i-} - \tilde{x}}{h_0} \\ \hat{m}_2(x_2) &:= \frac{\hat{T}_2(x_2)}{\hat{f}_2(x_2)}, \quad \hat{T}_2(x_2) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^n D_{2,K_2} \cdot \frac{X_{2-} - x_2}{h_2}, \quad \hat{f}_2(x_2) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^n K_2 \cdot \frac{X_{2-} - x_2}{h_2} \\ \hat{t}_i &:= 1\{f_X(\tilde{X}) \geq \tau_i\}, \quad \hat{f}_X(\tilde{x}) := \frac{1}{nb^q} \sum_{i=1}^n K_3 \cdot \frac{\tilde{X}_i - \tilde{x}}{b}. \end{aligned}$$

Their biases are estimated by

$$\begin{aligned} \hat{B}_0(\tilde{x}) &:= \frac{\hat{T}_{\hat{m}_0}(\tilde{x})}{\hat{f}_0(\tilde{x})} - \hat{m}_0(\tilde{x}), \quad \text{where} \quad \hat{T}_{\hat{m}_0}(\tilde{x}) := \frac{1}{nh_0^{p_0}} \sum_{i=1}^n \hat{m}_0(\tilde{Z}_i) \cdot \hat{t}_i \cdot K_0 \cdot \frac{Z_{i-} - \tilde{x}}{h_0} \\ \hat{B}_2(x_2) &:= \frac{\hat{T}_{\hat{m}_2}(x_2)}{\hat{f}_2(x_2)} - \hat{m}_2(x_2), \quad \text{where} \quad \hat{T}_{\hat{m}_2}(x_2) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^n \hat{m}_2(\tilde{X}_i) \cdot \hat{t}_i \cdot K_2 \cdot \frac{X_{2-} - x_2}{h_2} \end{aligned}$$

So, the bias-corrected estimators of the first step are

$$\tilde{m}_0(\tilde{x}) := \hat{m}_0(\tilde{x}) - \hat{B}_0(\tilde{x}) \quad \text{and} \quad \tilde{m}_2(x_2) := \hat{m}_2(x_2) - \hat{B}_2(x_2).$$

Denote

$$\begin{aligned} \tilde{W}(\beta) &:= (\beta_1^T X_1 + \beta_2^T \tilde{m}_2(X_2), x^e - \tilde{m}_0(\tilde{X}))^T, \quad \hat{W}(\beta) := (\beta_1^T X_1 + \beta_2^T \hat{m}_2(X_2), x^e - \hat{m}_0(\tilde{X}))^T, \\ \tilde{w}(\beta) &:= (\beta_1^T x_1 + \beta_2^T \tilde{m}_2(x_2), x^e - \tilde{m}_0(\tilde{z}))^T, \quad \text{and} \quad w(\beta) := (\beta_1^T x_1 + \beta_2^T \hat{m}_2(x_2), x^e - \hat{m}_0(\tilde{z}))^T \end{aligned}$$

For the estimators of  $tt_{W(\beta)}$  and  $f_{W(\beta)}$ ,

$$\begin{aligned} \hat{t}_{\tilde{W}(\beta)}(w) &:= \frac{\hat{T}_{\tilde{W}(\beta)}(w)}{\hat{f}_{\tilde{W}(\beta)}(w)}, \\ \hat{T}_{\tilde{W}(\beta)}(w) &:= \frac{1}{nh^d} \sum_{i=1}^n Y_i K \cdot \frac{\tilde{W}_i(\beta) - w}{h} \cdot \hat{t}_i, \quad \hat{f}_{\tilde{W}(\beta)}(w) := \frac{1}{nh^d} \sum_{i=1}^n K \cdot \frac{\tilde{W}_i(\beta) - w}{h} \cdot \hat{t}_i. \end{aligned}$$

For the bias of  $\hat{t}_{\tilde{W}(\beta)}$  and its bias-corrected version,

$$\begin{aligned} \hat{B}_{\tilde{W}(\beta)}(w) &:= \frac{\hat{T}_{\tilde{W}(\beta)}(w)}{\hat{f}_{\tilde{W}(\beta)}(w)}, \quad \text{with} \quad \hat{T}_{\hat{t}_{\tilde{W}(\beta)}}(w) := \frac{1}{nh^d} \sum_{i=1}^n \hat{t}_{\tilde{W}(\beta)} \cdot \hat{t}_i \cdot K \cdot \frac{\tilde{W}_i(\beta) - w}{h}, \\ \tilde{t}_{\tilde{W}(\beta)}(w) &:= \hat{t}_{\tilde{W}(\beta)}(w) - \hat{B}_{\tilde{W}(\beta)}(w) \end{aligned}$$

$$p_n := P \left( f(W(\beta_0)) \geq \tau_n \right), p_{n,0} := P \left( f_0(Z) \geq \tau_n \right), \\ p_{n,2} := P \left( f_2(X_2) \geq \tau_n \right), p_{n,3} := P \left( f_{\tilde{X}}(\tilde{X}) \geq \frac{3\tau_n}{2} \right)$$

We now list the Assumptions we will be using throughout this Appendix.

**Assumption A1 (IID).**  $\{Y_i, X_{1,i}, X_{2,i}, \tilde{X}_i, D_{1,i}, D_{2,i}\}_{i=1}^n$  is an IID sequence of bounded random variables defined over the probability space  $(\Omega, \mathcal{A}, P)$ .

**Assumption A2 (Smoothness).** (i) The mappings  $(w, \beta) \mapsto t_{W(\beta)}(w)$  and  $(w, \beta) \mapsto f_{W(\beta)}(w)$  belong to the class  $\mathbf{G}^r$ ; (ii) for all  $\beta \in \mathcal{B}$ ,  $t$  admits a density conditionally on the multiple index  $W(\beta)$  which is denoted as  $(\tilde{x}, w) \mapsto f_{\tilde{X}|W(\beta)}(\tilde{x}|w)$  and belongs to the class  $\mathbf{E}_{r_0 \vee r_2}^r$ ; (iii)  $m_0, f_0 \in \mathbf{G}^{r_0}$ ;  $m_2, f_2 \in \mathbf{G}^{r_2}$ ; (iv) the density  $f_{\tilde{X}}$  of  $\tilde{X}$  belongs to the class  $\mathbf{G}^{r_3}$ .

**Assumption A3 (Kernel).** (i)  $K \in \mathbf{K}_{d'}^r$ ; (ii)  $K_0 \in \mathbf{K}_{p_0}^{r_0}, K_2 \in \mathbf{K}_{p_2}^{r_2}$ ; (iii)  $K_{\tilde{X}} \in \mathbf{K}_{\dim(\tilde{X})}^{r_3}$ .

**Assumption A4 (Bandwidth rates).** (i) for  $j = 0, 2$ :  $\frac{\log n}{n h_j^{p_j+2s_j} \tau_n^{2s_j+2}} \rightarrow 0, \frac{h_j^{r_j}}{\tau_n} \rightarrow 0$ , for  $s_j = 1 + \underline{Ev}(p_j)/2$ ;  
 $\frac{(\log n)^2}{n h_j^{2p_j} \tau_n^2} \rightarrow 0, n \cdot h_j^{4r_j} \cdot \tau_n^{-4} \rightarrow 0$ ; (ii)  $\frac{(\log n)^2}{n h^{2d+4} \tau_n^8} \rightarrow 0, n \cdot h^{4r} \rightarrow 0$ ;

**Assumption A5 (Trimming).** (i)  $p_{n,0} = o(l_n^{-1}), p_{n,2} = o(l_n^{-1})$ , where  $l_n$  satisfies  $\frac{n^{1/4}}{h^{d+1} \tau_n^2 l_n} \rightarrow 0, \frac{n^{1/2}}{l_n} \rightarrow +\infty; \frac{h_j}{\tau_n} = o(1)$  for  $j = 0, 2$ , and  $\frac{h}{\tau_n} = o(1)$ .  
(ii)  $p_{n,3} = o(l_{n,3}^{-1})$  and the bandwidth  $b$  used for the estimation of  $f_{\tilde{X}}$  satisfies:  $n \cdot b^{q+5} \cdot \tau_n \rightarrow +\infty, \frac{\log n}{n \cdot b \tau_n^2} \rightarrow 0, \frac{b^{r_3}}{\tau_n} \rightarrow 0, \tau_n \cdot l_{n,3} \cdot b^q \rightarrow +\infty$ ;  
(iii) there exists constants  $\eta > 0$  small enough and  $N$  large enough, such that for all  $n$  larger than  $N$ ,  $\mathbf{1}\{f_{\tilde{X}}(\tilde{x}) \geq \tau_n\} \leq \sup_{\beta \in \mathcal{B}} \mathbf{1}\{f_{W(\beta)}(w) \geq \eta \tau_n\}, \mathbf{1}\{f_{\tilde{X}}(\tilde{x}) \geq \tau_n\} \leq \mathbf{1}\{f_{\tilde{X}}(\tilde{x}) \geq \eta \tau_n\}$ , and  $\mathbf{1}\{f_{\tilde{X}}(\tilde{x}) \geq \tau_n\} \leq \mathbf{1}\{f_2(x_2) \geq \eta \tau_n\}$ . For a large enough  $n$ , the set  $w : f_{W(\beta)}(w) \geq \frac{\eta}{2} \tau_n, z : f_0(z) \geq \frac{\eta}{2} \tau_n$ , and  $x_2 : f_2(x_2) \geq \frac{\eta}{2} \tau_n$  are convex.

**Assumption A6 (Pseudo-true value).** The mapping  $\beta \mapsto \mathbb{E}\{(Y - t_{W(\beta)}(W(\beta)))^2\}$  admits a unique minimum.

## A Asymptotic Expansion

Having specified the notation, I start by presenting some important objects that will be used throughout this appendix. Let  $\mathbf{F}$  be a space of real-valued functions defined over  $\tilde{X}$  and metricized by  $L_2(P)$ . So, for any  $f, g \in \mathbf{F}$ , the distance between  $f$  and  $g$  is measured by the (pseudo) norm  $\|f - g\|_{L_2(P)} = \left\{ \int |f - g|^2 dP(\tilde{x}) \right\}^{1/2}$ . For any two functions  $\tilde{x} \mapsto u(\tilde{x})$  and  $\tilde{x} \mapsto l(\tilde{x})$ , such that  $u \geq l$ , define the bracket  $[u, l] := \{f \in \mathbf{F} : l \leq f \leq u\}$ .

The bracket  $[u, l]$  has  $L_2(P)$ -size  $s$  if  $\|u - l\|_{L_2(P)} \leq s$ . Hence, define  $N_{[]} (s, F, L_2(P))$  to be the bracketing number of the semi-metric space  $(F, L_2(P))$ , i.e.  $N_{[]} (s, F, L_2(P))$  represents the minimum number of brackets of  $L_2(P)$  size  $s$  that covers the space  $F$ . Also, let  $J_{[]} (\delta, F, L_2(P)) := \int_0^\delta \frac{1}{s} \log N_{[]} (s, F, L_2(P)) ds$ . For a deeper treatment of these concepts, see [Kosorok \(2007\)](#), [Pollard \(1984\)](#), [van der Vaart & Wellner \(1996\)](#), and [van der Vaart \(1998\)](#). The lemma that follows is the main one used in the proofs.

**Lemma A.1.** ([van der Vaart \(1998\)](#)) Let  $F$  be a class of measurable functions  $f : \mathcal{X} \rightarrow \mathbb{R}$  such that  $Pf^2 < \delta^2$  for all  $f \in F$ , let  $F$  be the envelope function for  $F$ , and let  $a_F(\delta) := \delta / (1 \vee \log N_{[]} (\delta, F, L_2(P)))$ . Then

$$E\|G_n\|_F \leq J_{[]} (\delta, F, L_2(P)) + \sqrt{n} P F \{F > \sqrt{n} a_F(\delta)\}$$

up to a universal constant.

The following “lemmata” is necessary to prove the expansion of the empirical process at the basis of the statistic.

**Lemma A.2.** Let Assumption A5 hold. Let  $(y, \tilde{x}) \mapsto \hat{g}_{n,t}(y, \tilde{x})$  be a sequence of random functions with  $t \in \mathcal{T}$ , such that  $\sup_{t \in \mathcal{T}} \|\hat{g}_{n,t}\|_\infty = O_P(1)$  and  $\sup_{t \in \mathcal{T}} \|\hat{g}_{n,t_n}\|_\infty = O_P(1)$ . The following result hold:

$$(i) \sqrt{n} \cdot P_n |t_n - 1| = o_P(1);$$

$$(ii) \sqrt{n} \cdot P_n \hat{g}_{n,t} - \sqrt{n} \cdot P_n \hat{g}_{n,t_n} = o_P(1) \text{ uniformly in } t \in \mathcal{T};$$

Let  $z \mapsto \psi_t(z)$  be a function such that  $\sup_{t \in \mathcal{T}} \|\psi_t\|_\infty$  is finite. Then,

$$(iii) \sqrt{n} \cdot P_n \hat{g}_{n,t} \cdot \int K_0(u) \cdot \psi_t(Z + uh_0) \cdot t_{n,Z}^{(\eta/2)}(Z + uh_0) du = \sqrt{n} \cdot P_n \hat{g}_{n,t_n} \cdot \int K_0(u) \cdot \psi_t(Z + uh_0) du + o_P(1) \text{ uniformly in } t \in \mathcal{T};$$

Let  $w \mapsto \tilde{\psi}_t(w)$  be a function such that  $\sup_{t \in \mathcal{T}} \|\tilde{\psi}_t\|_\infty$  is finite. Then,

$$(iv) \sqrt{n} \cdot P_n \hat{g}_{n,t} \cdot \int K(u) \cdot \tilde{\psi}_t(W(\beta_0) + uh) \cdot t_{n,W(\beta_0)}^{(\eta/2)}(W(\beta_0) + uh) du = \sqrt{n} \cdot P_n \hat{g}_{n,t_n} \cdot \int K(u) \cdot \tilde{\psi}_t(W(\beta_0) + uh) \cdot t_{n,W(\beta_0)}^{(\eta/2)}(W(\beta_0) + uh) du \text{ uniformly in } t \in \mathcal{T}.$$

If  $\sup_{\tilde{x}} |w(\tilde{\beta}) - w(\beta_0)| \cdot t_{\tilde{x}}(\tilde{x})| = O(\tilde{x})$ , the above result also holds by replacing  $W(\beta)$  with  $W(\tilde{\beta})$ .

*Proof.* (i)  $\sqrt{n} \cdot P |t_n - 1| = \sqrt{n} \cdot P \{f(\tilde{X}) < \tau_n\} = o(1)$ , by Assumption A5. So, conclude by Markov's inequality.

(ii) Notice first that  $t_{\frac{1}{n}} t - t_{\frac{1}{n}} t - \frac{2}{n} t = \frac{1}{n} t - \hat{t}_n(t - \frac{1}{n} t)$ , so

$$\sqrt{n} \cdot P_n |g_{n,t} \cdot (t_n - t_n)| \leq \|g_{n,t} \cdot (t_n - t_n)\|_\infty \sqrt{n} \cdot P_n |t_n - t_n| = O_P(\sqrt{n} \cdot P_n |t_n - t_n|).$$

Define

$$B_n^{(C)} := \|f_{\tilde{X}} - f_{\tilde{X}}\|_\infty \leq C \cdot d_{3,n}, \text{ with } d_{3,n} = \frac{\log n}{n b^q} + b^3$$

By [Li & Racine \(2006\)](#),  $\|f_{\tilde{X}} - f_{\tilde{X}}\|_\infty = O_P(d_{3,n})$ , so  $P(B_n^{(C)})$  can be made arbitrarily close to one, asymptotically, by choosing  $C$  large enough. By Assumption A5,  $d_{3,n}/\tau_n = o(1)$ . Notice that if  $B_n^{(C)}$  holds,  $f_{\tilde{X}}(\tilde{x}) \geq \tau \cdot \frac{3}{2}$ ,

and  $Cd_{3,n}/\tau_n < 1/2$ , then  $t_n(\tilde{x}) = 1$  and  $f_{\tilde{X}}(\tilde{x}) > \frac{f_{\tilde{X}}(\tilde{x})}{\tau_n} - C \frac{d_{3,n}}{\tau_n} \cdot \tau_n > \tau_n$ . So, whenever  $B_n^{(C)}$  holds and  $Cd_{3,n}/\tau_n < 1/2$ , it must be that

$$|\hat{t}_n(\tilde{x}) - t_n(\tilde{x})| \leq \frac{f_{\tilde{X}}(\tilde{x})}{\tau_n} < \frac{3}{2}\tau_n.$$

Since  $\sqrt{n} \cdot P_{\tilde{X}}(f_{\tilde{X}}(\tilde{X}) < \frac{3}{2}\tau_n) = o(1)$ , conclude by Markov's inequality.

(iii) By Assumption A5, for  $n$  large enough  $t_n = t_{n,Z}^{(\eta)}$ , and hence

$$\begin{aligned} & \sqrt{n} \cdot P_{n, \hat{g}_{n,t_n}} \int_{t_n}^{\infty} K_0(u) \cdot \psi_t(Z + uh_0) \cdot t_{n,Z}^{(\eta/2)}(Z + uh_0) du = \\ & \sqrt{n} \cdot P_{n, \hat{g}_{n,t_n}} \int_{t_n}^{t_{n,Z}^{(\eta)}} K_0(u) \cdot \psi_t(Z + uh_0) \cdot t_{n,Z}^{(\eta/2)}(Z + uh_0) du. \end{aligned} \quad (A.1)$$

By a first-order Taylor expansion,  $f_Z(z + uh_0) = f_Z(z) + \partial f_Z(z) \cdot (uh_0)$ , with  $\tau \in [z, z + uh_0]$  and  $|\partial f_Z(\tau) \cdot uh_0| \leq Ch_0$  for all  $u$  such that  $K_0(u) \neq 0$ . Hence, since  $\frac{h_0}{\tau_n} = o(1)$ , for a large  $n$ ,

$$\begin{aligned} t_{n,Z}^{(\eta/2)}(z + uh_0) &= f_Z(z) \geq \frac{1}{2} - \frac{\partial f_Z(\tau)}{\tau_n} \cdot \frac{uh_0}{\tau_n} \geq \\ & f_Z(z) \geq \frac{\eta\tau_n}{2} \cdot 1 + \frac{1}{3} \geq t_{n,Z}^{(\eta)}(z) \end{aligned}$$

uniformly in  $z$  and for all  $u$  such that  $K_0(u) \neq 0$ . The result then follows from the above display and Eq. (A1).

The proof of Point (iv) is very similar to the proof of Point (iii), so it is omitted.  $\square$

**Lemma A.3.** Let Assumptions A1-A5 hold. If  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$  and  $H_0$  holds, uniformly over  $\{\tilde{x} : f(\tilde{x}) \geq \tau_n\}$

- (i)  $\hat{T}_{\tilde{W}(\hat{\beta})}(w(\hat{\beta})) = \hat{T}_{\tilde{W}(\hat{\beta})}(w(\beta_0)) + \partial T_{\tilde{W}(\hat{\beta})}(w(\beta_0)) \cdot (\tilde{w}(\hat{\beta}) - w(\beta_0)) + o_P(n^{-1/2})$
- (ii)  $\hat{f}_{\tilde{W}(\hat{\beta})}(w(\hat{\beta})) = \hat{f}_{\tilde{W}(\hat{\beta})}(w(\beta_0)) + \partial f_{\tilde{W}(\hat{\beta})}(w(\beta_0)) \cdot (\tilde{w}(\hat{\beta}) - w(\beta_0)) + o_P(n^{-1/2})$
- (iii)  $\int K(u) \cdot (\psi_{f_{\tilde{W}(\hat{\beta})}}(w(\hat{\beta}) + uh) du = \int K(u) \cdot (\psi_{f_{\tilde{W}(\beta_0)}}(w(\beta_0) + uh) du + \partial(\psi_{f_{\tilde{W}(\beta_0)}})(w(\beta_0)) \cdot (\tilde{w}(\hat{\beta}) - w(\beta_0)) + o_P(n^{-1/2})$

The above results also hold with  $\tilde{w}(\hat{\beta})$  replaced by  $w(\hat{\beta})$  and  $\tilde{W}(\hat{\beta})$  replaced by  $\hat{W}(\hat{\beta})$ .

*Proof.* The proof is provided in a supplementary material. (i) By a Mean-Value expansion,

$$\hat{T}_{\tilde{W}(\hat{\beta})}(w(\hat{\beta})) = \hat{T}_{\tilde{W}(\hat{\beta})}(w(\beta_0)) + \partial \hat{T}_{\tilde{W}(\hat{\beta})}(\overline{w(\beta)}) \cdot (\tilde{w}(\hat{\beta}) - w(\beta_0)), \text{ with } \overline{w(\beta)} \in [\tilde{w}(\hat{\beta}), w(\beta_0)]. \quad (A.2)$$

Define the event  $B_n := \{\sup_{\tilde{x}} t_n(\tilde{x}) \cdot |\tilde{w}(\hat{\beta}) - w(\beta_0)| \leq C \cdot d_n/\tau_n\}$ . By Lemma D.1,  $P(B_n)$  has a probability arbitrarily close to one asymptotically by choosing  $C$  large enough. For a fixed  $C$ , over the set  $B_n$ , whenever  $t_n(\tilde{x}) = 1$ , by a Mean-Value expansion,

$$f_{\tilde{W}(\beta_0)}(\overline{w(\beta)}) = f_{\tilde{W}(\beta_0)}(w(\beta_0)) + \partial f_{\tilde{W}(\beta_0)}(\overline{w(\beta)}) \cdot (\overline{w(\beta)} - w(\beta_0)),$$

with  $\overline{w(\beta)} \in [\overline{w(\beta)}, w(\beta_0)]$ , and

$$|\partial f_{W(\beta_0)}(\overline{w(\beta)}) \cdot (\overline{w(\beta)} - w(\beta))|_{t_n(x^\sim)} \leq \|\partial f_{W(\beta_0)}\|_\infty \cdot C \cdot d_n/\tau_n,$$

where  $\frac{d_n}{\tau_n} = o(1)$ . Hence, by choosing a  $\theta \in (\frac{2}{\eta}, \frac{3}{\eta})$  over the set  $B$ , for each  $n$  sufficiently large

$$\begin{aligned} \{f_{\tilde{X}}(x^\sim) \geq \tau_n\} \cdot f_{W(\beta)}(\overline{w(\beta)}) &\geq \frac{\tau_n}{\theta} = \frac{\tau_n}{\theta} \cdot [1 - \theta \cdot \partial f_{W(\beta)}(\overline{w(\beta)}) \cdot (w(\beta) - \overline{w(\beta)})/\tau_n] \geq \\ \{f_{\tilde{X}}(x^\sim) \geq \tau_n\} \cdot f_{W(\beta)}(w(\beta_0)) &\geq \frac{\tau_n}{\theta} \cdot [1 + \theta \cdot \|\partial f_{W(\beta)}\|_\infty \cdot C \cdot \frac{d_n}{\tau_n}] \geq \\ \{f_{\tilde{X}}(x^\sim) \geq \tau_n\} \cdot f_{W(\beta)}(w(\beta_0)) &\geq \frac{\tau_n}{\theta} \cdot 2 \geq \\ \{f_{\tilde{X}}(x^\sim) \geq \tau_n\} \cdot \{f_{W(\beta_0)}(w(\beta_0)) \geq \eta \tau_n\} &\geq \\ \{f_{\tilde{X}}(x^\sim) \geq \tau_n\} & \end{aligned}$$

for all  $x^\sim$ , where the last inequality follows from, Assumption A5. Accordingly, by Lemma D.3, whenever the event  $B_n$  holds true

$$\begin{aligned} \sup_{\tilde{x}} |\partial \hat{T}_{\tilde{W}(\hat{\beta})}(\overline{w(\beta)}) - \partial T_{W(\beta)}(\overline{w(\beta)})| \cdot t_n(x^\sim) &\leq \\ \sup_w |\partial \hat{T}_{\tilde{W}(\hat{\beta})}(w) - \partial T_{W(\beta)}(w)| \cdot \{f_{W(\beta)}(w) \geq \tau_n/\theta\} &\leq \\ \sup_w |\partial \hat{T}_{\tilde{W}(\hat{\beta})}(w) - \partial T_{W(\beta)}(w)| \cdot \{f_{W(\beta)}(w) \geq \eta \tau_n/3\} &= o_P(n^{-1/4}). \end{aligned} \quad (\text{A.3})$$

Finally, notice that by a Mean-Value Theorem, uniformly in  $x^\sim$ ,

$$|\partial T_{W(\beta_0)}(w(\beta)) - \partial T_{W(\beta_0)}(w(\beta_0))| \cdot t_n(x^\sim) \leq C \cdot |w(\beta) - w(\beta_0)| \cdot t_n(x^\sim) = o_P(n^{-1/4}), \quad (\text{A.4})$$

where the last equality follows from the definition of  $\overline{w(\beta)}$ , Lemma D.1, and  $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$ . Notice now that by Lemma D.1 and  $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$ ,

$$\sup_{\tilde{x}} |(\tilde{w}(\hat{\beta}) - w(\beta_0))| \cdot t_n(x^\sim) = o_P(n^{-1/4}). \quad (\text{A.5})$$

Conclude by Eq. (A.2), (A.3), (A.4), and (A.5). The proof of point (ii) proceeds along the same lines.

(iii) The proof proceeds along the same lines as above, by using the fact that  $\sup_{t \in T} |\partial(\iota f_{W(\beta_0)})(w_1) - \partial(\iota f_{W(\beta_0)})(w_2)| \leq C|w_1 - w_2|$ ,  $n^{1/4}h^r = o(1)$  by Assumption A4, and  $K(u) \cdot \partial(\iota f_{W(\beta_0)})(w(\beta_0) + uh) du = \partial(\iota f_{W(\beta_0)})(w(\beta_0)) + O(h^r)$  uniformly over the set  $\{x^\sim : f(x^\sim) \geq \tau_n\}$  and  $\mathbb{T}$ , by an usual Taylor expansion.  $\square$

**Lemma A.4.** Let  $\Phi := \{x^\sim \mapsto \phi(x^\sim) : t \in T\}$  with  $\phi$  fixed functions satisfying  $\sup_{x^\sim \in \text{Supp } X} |\phi_{t_1}(x^\sim) - \phi_{t_2}(x^\sim)| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in \mathbb{T}$ . Let  $\mathbf{C}$  be a class of functions such that  $\log N(s, \mathbf{C}, \|\cdot\|_\infty) \leq C \cdot s^{-v}$ . Then  $\log N[\cdot](s, \mathbf{C} \cdot \Phi, \|\cdot\|_{2,P}) \leq Cs^{-v}$ , with  $v \in (0, 2)$ .



*Proof.* Since  $\mathcal{T}$  is a compact set,  $N(s, \mathcal{T}, \|\cdot\|) \leq Cs^{-\dim(\mathcal{T})}$ , where  $\|\cdot\|$  denotes the euclidean norm over  $\mathbb{R}^{\dim(\mathcal{T})}$ . By definition of covering number and the Lipschitz property of the class  $\Phi$ ,  $N(Cs, \Phi, \|\cdot\|_\infty) \leq N(s, \mathcal{T}, \|\cdot\|)$ . Consider now the  $\delta$ -covers  $\mathbf{A}_C := \{g_I : I = 1, \dots, N(\delta, C, \|\cdot\|_\infty)\}$  and  $\mathbf{A}_\Phi := \{\varphi_J : J = 1, \dots, N(\delta, \Phi, \|\cdot\|_\infty)\}$ . For the generic element of  $C \cdot \Phi$ , say  $g \cdot \varphi$ , with  $g \in C$  and  $\varphi \in \Phi$ ,  $\|g - g_I\|_\infty < \delta$  and  $\|\varphi - \varphi_J\|_\infty < \delta$ , for some  $g_I \in \mathbf{A}_C$  and  $\varphi_J \in \mathbf{A}_\Phi$ . Hence,  $\|g \cdot \varphi - g_I \cdot \varphi_J\|_\infty \leq \|(g - g_I) \cdot \varphi\|_\infty + \|g_I \cdot (\varphi - \varphi_J)\|_\infty < C\delta$ . The collection  $\{g_I \cdot \varphi_J : g_I \in \mathbf{A}_C, \varphi_J \in \mathbf{A}_\Phi\}$  forms a  $C\delta$  cover for the class  $\Phi \cdot C$ , and hence

$$N(C\delta, \Phi \cdot C, \|\cdot\|_\infty) \leq \#\mathbf{A}_C \cdot \#\mathbf{A}_\Phi \leq N(\delta, C, \|\cdot\|_\infty) \cdot N(\delta, \Phi, \|\cdot\|_\infty).$$

Accordingly,  $\log N(C\delta, \Phi \cdot C, \|\cdot\|_\infty) \leq C \cdot \delta^{-v}$ . By proceeding as in Corollary 2.7.1 in [van der Vaart & Wellner \(1996\)](#),  $N_{[]}(\delta, \Phi \cdot C, L_2(P)) \leq N(\delta/2, \Phi \cdot C, \|\cdot\|_\infty)$ . Conclude from this last inequality.  $\square$

We now introduce a class of Holder continuous functions that will be used throughout this Appendix. For any function  $g : \mathcal{W} \rightarrow \mathbb{R}$ , and any vector  $s = (s_1, \dots, s_d)$ , define the differential operator  $\partial^s$  as

$$\partial^s g(w) = \frac{\partial^{|s|}}{\partial^{s_1} w_1 \dots \partial^{s_d} w_d} g(w),$$

where  $|s| = s_1 + \dots + s_d$ . Also, define the following set

$$\begin{aligned} \mathcal{W}_n &:= \{w : f_{W(\beta)}(w) > \frac{\eta}{2n}\}, \\ \mathcal{C}(\mathcal{W}_n) &:= \{g : \mathcal{W}_n \rightarrow \mathbb{R} \text{ s.t. } \|\partial^s g\|_{\mathcal{W}_n, \infty} \leq M \text{ for all } s \text{ with } |s| \leq \frac{Ev(d)}{2} + 1\}, \end{aligned} \quad (\text{A.6})$$

where  $Ev(d)$  denotes the largest even integer weakly smaller than  $d$  and  $\|f\|_{\mathcal{Z}_n, \infty} := \sup_{z \in \mathcal{Z}_n} |f(z)|$  for any function  $f$ . The class  $\mathcal{C}(\mathcal{W}_n)$  defined in the above display is a subset of the Holder continuous class of functions defined in [van der Vaart & Wellner \(1996\)](#), pages 154-155. Hence, from the compactness of the support of  $W(\beta_0)$ , the definition of  $\mathcal{W}_n$ , and Theorem 2.7.1 in [van der Vaart & Wellner \(1996\)](#) it holds that

$$\log N(s, \mathcal{C}(\mathcal{W}_n), \|\cdot\|_{\mathcal{W}_n, \infty}) \leq C \cdot s_v^{-v} \text{ with } v \in (0, 2), \quad (\text{A.7})$$

where the constant  $C$  does not depend on neither  $n$  nor  $s$ .

**Lemma A.5.** Assume that  $H_0$  holds and that  $\sqrt{n}(\hat{\beta} - \beta_0) = O_P(1)$ . Under Assumption A1-A4, uniformly in  $t \in \mathcal{T}$

- (i)  $G_n(\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\tilde{W}(\hat{\beta})}) \varphi_t t_n = o_p(1)$ ;
- (ii)  $G_n \varepsilon \hat{f}_{\tilde{W}(\hat{\beta})} \varphi_t t_n = o_p(1)$ ;
- (iii)  $G_n(\hat{T}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})}) \varphi_t t_n = o_p(1)$ ;

The same result holds when replacing  $\tilde{W}(\hat{\beta})$  with  $\hat{W}(\hat{\beta})$ .

*Proof.* (i) Define  $\Phi := \{\tilde{x} \mapsto \varphi(\tilde{x}t) : t \in \mathcal{T}\}$ , and for simplicity let  $\mathcal{C}_n$  denote the class  $\mathcal{C}(\mathcal{W}_n)$  defined in Eq. (A.6). From Lemma D.4, the mapping  $\tilde{x} \mapsto (\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\tilde{W}(\hat{\beta})})(w(\beta_0)) \cdot \varphi(\tilde{x}t) \cdot t_n(\tilde{x})$  belongs to the class

$t_n \cdot C_n \cdot \Phi = \{x \mapsto t_n(x) \cdot g(x) \cdot \psi(x) : g \in C_n \text{ and } \psi \in \Phi\}$  for all  $t \in T$ , with a probability approaching one. Fix  $\delta > 0$  arbitrarily small, and consider a  $\delta$ -cover for  $C_n$ , say  $A_n := \{g_j : j = 1, \dots, N(\delta, C_n, \|\cdot\|_{W_n, \infty})\}$ . For any element of  $t_n C_n$ , say  $t_n \cdot g$ , it must be that  $g \in C_n$ , and hence, for some  $g_j \in A_n$ ,  $\|g - g_j\|_{W_n, \infty} < \delta$ . Hence from Assumption A5,  $\|(g - g_j) \cdot t_n\|_\infty \leq \|g - g_j\|_{W_n, \infty} < \delta$ . Accordingly,

$$N(\delta, t_n C_n, \|\cdot\|_\infty) \leq \#A_n \leq N(\delta, C_n, \|\cdot\|_{W_n, \infty}).$$

From Lemma A.4 and Eq. (A.7),

$$\log N_{[]}(\varepsilon, t_n \cdot C_n \cdot \Phi, L_2(P)) \leq C\varepsilon^{-\nu} \quad \text{with } \nu \in (0, 2).$$

Furthermore, Lemma D.3 ensures that  $\sup_{t \in T} \|\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta)} \hat{f}_{\tilde{W}(\hat{\beta})} \cdot \varphi_t \cdot t_n\|_{2,P} = o_P(1)$ . Accordingly, for any  $\delta > 0$  arbitrarily small, the event  $B_n^{(\delta)} := \{\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta)} \hat{f}_{\tilde{W}(\hat{\beta})} \cdot \varphi_t \cdot t_n \in (t \cdot C_n \cdot \Phi)^\delta \text{ for all } t \in T\}$  has a probability approaching one, where  $(t_n \cdot C_n \cdot \Phi)^\delta := \{f \in t_n \cdot C_n \cdot \Phi : \|f\|_{2,P} < \delta\}$ . Over the set  $B_n^{(\delta)}$ ,

$$\sup_{t \in T} \|\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta)} \hat{f}_{\tilde{W}(\hat{\beta})} \cdot \varphi_t \cdot t_n\| \leq \|G_n\|_{(t_n \cdot C_n \cdot \Phi)^\delta}.$$

Now, by Lemma A.1,

$$\mathbb{E} \|G_n\|_{(t_n \cdot C_n \cdot \Phi)^\delta} \leq J_{[]}(\delta, (t_n \cdot C_n \cdot \Phi), L_2(P)) + \sqrt{n} C \{C > a_{(t_n \cdot C_n \cdot \Phi)^\delta}(\delta) \cdot \sqrt{n}\}.$$

Since  $(t_n \cdot C_n \cdot \Phi)^\delta \subset t_n \cdot C_n \cdot \Phi$  for  $\delta < 1$ ,

$$\log N_{[]}(\varepsilon, (t_n \cdot C_n \cdot \Phi)^\delta, L_2(P)) \leq \log N_{[]}(\varepsilon, t_n \cdot C_n \cdot \Phi, L_2(P)), \text{ with } J_{[]}(\varepsilon, t_n \cdot C_n \cdot \Phi, L_2(P)) < \infty.$$

Thus for  $\delta \rightarrow 0$ ,

$$J_{[]}(\delta, (t_n \cdot C_n \cdot \Phi)^\delta, L_2(P)) \leq J_{[]}(\delta, t_n \cdot C_n \cdot \Phi, L_2(P)) \rightarrow 0.$$

For any fixed  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \sqrt{n} C \{C > a_{(t_n \cdot C_n \cdot \Phi)^\delta}(\delta) \cdot \sqrt{n}\} \leq \limsup_{n \rightarrow \infty} a_{(t_n \cdot C_n \cdot \Phi)^\delta}(\delta)^{-1} C \{C > a_{(t_n \cdot C_n \cdot \Phi)^\delta}(\delta) \cdot \sqrt{n}\} = 0.$$

So, conclude by Markov's inequality.

The proofs for point (ii) and (iii) follow from the same arguments. □

**Lemma A.6.** Under Assumptions A1-A5, uniformly over  $T$ ,

- (i)  $G_n \varepsilon \int K(u) \cdot (\iota_{t f_{W(\beta_0)}})(W(\beta_0) + uh) du = G_n \varepsilon \cdot (\iota_{t f_{W(\beta_0)}})(W(\beta_0)) + o_P(1);$
- (ii)  $G_n(D_0 - m_0) \int K(u) \cdot \phi_t(Z + uh_0) du = G_n(D_0 - m_0) \phi_t + o_P(1);$

*Proof.* (i) Define  $g_{n,t}(\varepsilon, \tilde{x}) := \varepsilon \int K(u) \cdot (\iota_{if_V(\beta_0)})(w(\beta_0) + uh) du - \varepsilon \cdot (\iota_{if_V(\beta_0)})(w(\beta_0))$ . Since from Assumption A2,  $\sup_w |\iota_{t_1}(w) - \iota_{t_2}(w)| \leq C|t_1 - t_2|$ , by the boundedness of the functions involved,  $|g_{n,t_1}(\varepsilon, \tilde{x}) - g_{n,t_2}(\varepsilon, \tilde{x})| \leq C|t_1 - t_2|$ . Accordingly, from Lemma A.4, the class  $G_n := \{g_{n,t} : t \in T\}$  is such that  $\log N_{[\cdot]}(\delta, G_n, \|\cdot\|_\infty) \leq C\delta^{-v}$ , with  $v \in (0, 2)$ . To show the  $L_2$ -convergence, notice that by a Lebesgue Dominated Convergence argument (LDC, now henceforth),  $\varepsilon \int K(u) \cdot (\iota_{if_V(\beta_0)})(v(\beta_0) + uh) du \rightarrow \varepsilon(\iota_{if_V(\beta_0)})(v(\beta_0))$  for any fixed  $(t, \tilde{x}, \varepsilon)$ . Hence, by applying again a LDC argument,  $\|g_{n,t}\|_{2,P} \rightarrow 0$  for any fixed  $t \in T$ . To make such a convergence uniform over  $T$ , fix  $\delta > 0$  arbitrarily small, and choose a  $\delta$ -cover for  $T$ , say  $A := \{t_j : j = 1, \dots, N(\delta, T, \|\cdot\|)\}$ , where  $\|\cdot\|$  denotes the Euclidean norm. For  $n$  large enough,  $\|g_{n,t_j}\|_{2,P}^2 < \delta$  for all  $t_j \in A$ . So, for any  $t \in T$ ,

$$\|g_{n,t}\|_{2,P}^2 \leq \|g_{n,t}\|_{2,P}^2 - \|g_{n,t_j}\|_{2,P}^2 + \|g_{n,t_j}\|_{2,P}^2 \leq C|t - t_j| + \delta^2 \leq C\delta^2 + \delta^2.$$

Conclude that  $\sup_{t \in T} \|g_{n,t}\|_\infty = o(1)$ . From this point, proceed as in Lemma A.5 to conclude. The proof of point (ii) proceeds along the same lines.  $\square$

Define the class

$$C(Z_n) := \left\{ z \mapsto g(z) : \|\partial^s g\|_{Z_n, \infty} \leq M \text{ for all } s \text{ such that } |s| \leq \frac{Ev(p_0)}{2} + 1 \right\} \text{ with}$$

$$Z_n := \left\{ z : f_0(z) \geq \frac{\eta_{\varepsilon_n}}{2} \right\},$$

where  $\|f\|_{Z_n, \infty} := \sup_{z \in Z_n} |f(z)|$  for any function  $f$ . Similarly to  $C(W_n)$ , the class  $C(Z_n)$  defined in the above display is a subset of the Holder continuous class of functions defined in [van der Vaart & Wellner \(1996\)](#), pages 154-155. From the compactness of the support of  $Z$ , the definition of  $Z_n$ , and Theorem 2.7.1 in [van der Vaart & Wellner \(1996\)](#) it holds that

$$\log N(s, C(Z_n), \|\cdot\|_{Z_n, \infty}) \leq C \cdot s^{-v} \text{ with } v \in (0, 2), \quad (\text{A.8})$$

where the constant  $C$  does not depend on neither  $n$  nor  $s$ .

**Lemma A.7.** Let  $(\tilde{x}, t) \mapsto \phi_t(\tilde{x})$  be a mapping such that  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in T$ , and  $\sup_{t \in T, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ . Under Assumption A1-A5, uniformly over  $T$ ,

$$G_n(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t = o_P(1), \text{ and } G_n(\hat{T}_{\hat{m}_0} \hat{f}_0 - m_0) \cdot t_n \cdot \phi_t = o_P(1).$$

*Proof.* Define the class  $Y := \{\tilde{x} \mapsto \phi_t(\tilde{x}) : t \in T\}$ . For simplicity denote with  $C_Y$  the class  $C(Y)$ . By the same reasoning as in the proof of Lemma A.5,

$$N(\delta, t_n C_n, \|\cdot\|_\infty) \leq \#A_n \leq N(\delta, C_n, \|\cdot\|_{Z_n, \infty}).$$

Define

$$t_n \cdot (C_n - m_0) := \{g - t_n \cdot m_0 : g \in t_n \cdot C_n\}.$$

Since  $m_0 t_n$  is a fixed function,  $N(\delta, t_n \cdot (C_n - m_0), \|\cdot\|_\infty) \leq N(\delta, t_n C_n, \|\cdot\|_\infty)$ . From Lemma A.4, the feature of  $(t, \tilde{x}) \mapsto \phi_t(\tilde{x})$ , the entropy bound just derived, and Eq. (A.8),

$$\log N_{[]}(\delta, t_n \cdot (C_n - m_0) \cdot Y, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v}.$$

Now, from Lemma D.1,  $P((\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t \in (C_n(Z_n) - m_0 t_n) \cdot \Phi) \rightarrow 1$ , and  $\sup_{t \in \mathbb{T}} \|(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t\|_{2,P} = o_P(1)$ . From this point onwards, proceed in the same way as in Lemma A.5 to obtain that  $G_n(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t = o_P(1)$ . The remaining result is obtained by the same steps.  $\square$

**Lemma A.8.** *Let  $(\tilde{x}, t) \mapsto \phi_t(\tilde{x})$  be a mapping such that  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in \mathbb{T}$ , and  $\sup_{t \in \mathbb{T}, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ . Under Assumption A1-A5, uniformly over  $\mathbb{T}$*

$$\begin{aligned} \sqrt{n} \cdot P_n \phi_t \cdot \hat{B}_0 \cdot t_n &= -\sqrt{n} \cdot P_n(D_0 - m_0) \cdot E\{\phi_t(\tilde{X}) | Z\} + \\ &\quad \sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot E\{\phi_t(\tilde{X}) | Z\} + o_P(1). \end{aligned}$$

The same type of result holds for  $\hat{B}_2$ .

*Proof.* By definition of  $\hat{B}_0$  and  $\hat{m}_0$ ,

$$\begin{aligned} \sqrt{n} \cdot P_n \phi_t \cdot \hat{B}_0 \cdot t_n &= G_n(\hat{T}_{\hat{m}_0} f_{\hat{f}_0} - m_0) \cdot t_n \cdot \phi_t \\ &\quad - G_n(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t + \sqrt{n} \cdot P(\hat{T}_{\hat{m}_0} f_{\hat{f}_0} - \hat{m}_0) \cdot t_n \cdot \phi_t. \end{aligned}$$

By Lemma A.7, uniformly over  $\mathbb{T}$ ,  $G_n(\hat{T}_{\hat{m}_0} f_{\hat{f}_0} - m_0) \cdot t_n \cdot \phi_t = o_P(1)$  and  $G_n(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t = o_P(1)$ . From Lemma D.1, Assumption A5, and Lemma A.2, uniformly in  $t \in \mathbb{T}$

$$\sqrt{n} \cdot P(\hat{T}_{\hat{m}_0} f_{\hat{f}_0} - \hat{m}_0) \cdot t_n \cdot \phi_t = \sqrt{n} \int (\hat{T}_{\hat{m}_0}(z) - \hat{T}(z)) \cdot t_{n,Z}^{(\eta/2)}(z) \cdot E\{\phi_t(\tilde{X}) | Z = z\} dz + o_P(1).$$

By the classical “change of variable” and Lemma A.2, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} &\sqrt{n} \int (\hat{T}_{\hat{m}_0}(z) - \hat{T}(z)) \cdot t_{n,Z}^{(\eta/2)}(z) \cdot E\{\phi_t(\tilde{X}) | Z = z\} dz = \\ &\quad -\sqrt{n} \cdot P_n(D_0 - m_0) \cdot \int K_0(u) \cdot \tilde{\phi}_t(z + u h_0) du + \\ &\quad \sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \int K_0(u) \cdot \tilde{\phi}_t(Z + u h_0) du + o_P(1), \end{aligned}$$

where  $\tilde{\phi}_t(z) := E\{\phi_t(\tilde{X}) | Z = z\}$ . Since  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in \mathbb{T}$ , and  $\sup_{t \in \mathbb{T}, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ , Lemma A.6 ensures that

$$\sqrt{n} \cdot P_n(D_0 - m_0) \cdot \int K_0(u) \cdot \tilde{\phi}_t(z + u h_0) du = \sqrt{n} \cdot P_n(D_0 - m_0) \cdot E\{\phi_t(\tilde{X}) | Z\} + o_P(1)$$

uniformly over  $\mathbb{T}$ . Finally, by Lemma D.1 and Assumption A4,

$$\sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \int K_0(u) \cdot \tilde{\phi}_t(Z + u h_0) du = \sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \tilde{\phi}_t(Z) + o_P(1),$$

uniformly over  $\mathbf{T}$ . Conclude by putting together the last three displays.  $\square$

**Lemma A.9.** Under Assumptions A1-A6, uniformly over  $\mathbf{T}$ ,

$$\sqrt{n} \cdot \mathbf{P}_n(\tilde{m}_0 - m_0) \cdot t_n \cdot \phi_t = -\sqrt{n} \cdot \mathbf{P}_n(D_0 - m_0) \cdot \mathbf{E}\{\phi_t(X^\sim) | Z\} + o_P(1).$$

The same type of expansion also holds for  $\tilde{m}_2$ .

*Proof.* From Lemma A.7, Lemma A.2, and the Law of Iterated Expectations,

$$\sqrt{n} \cdot \mathbf{P}_n(\tilde{m}_0 - m_0) \cdot t_n \cdot \phi_t = \sqrt{n} \cdot \mathbf{P}(\hat{m}_0 - m_0) \cdot t_n \cdot \mathbf{E}\{\phi_t(X^\sim) | Z\} - \sqrt{n} \cdot \mathbf{P}_n \phi_t \cdot \hat{B}_0 \cdot t_n + o_P(1),$$

uniformly in  $t \in \mathbf{T}$ . By this expansion and Lemma A.8,

$$\begin{aligned} \sqrt{n} \cdot \mathbf{P}_n(\tilde{m}_0 - m_0) \cdot t_n \cdot \phi_t &= \sqrt{n} \cdot \mathbf{P}_n(D_0 - m_0) \cdot \mathbf{E}\{\phi_t(X^\sim) | Z\} \\ &\quad - \sqrt{n} \cdot (\mathbf{P}_n - \mathbf{P})(\hat{m}_0 - m_0) \cdot t_n \cdot \mathbf{E}\{\phi_t(X^\sim) | Z\} + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ . Hence, conclude by Lemma A.7  $\square$

The lemmas that follow are simple applications of the previous ones.

**Lemma A.10.** Let  $H_0$  hold and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = o_P(1)$ . Under Assumptions A1-A6, uniformly in  $t \in \mathbf{T}$ ,

- (i)  $\mathbf{G}_{n\mathcal{E}} \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot q \cdot t_n = \mathbf{G}_{n\mathcal{E}} f_{W(\beta_0)}(W(\beta_0)) \cdot \varphi_t + o_P(1);$
- (ii)  $\sqrt{n} \cdot \mathbf{P}_n \varepsilon t_n \int K(u) \cdot (\iota_t f_{W(\beta_0)}(\tilde{W}(\hat{\beta}) + uh) du = \sqrt{n} \cdot \mathbf{P}_n \varepsilon \cdot (\iota_t f_{W(\beta_0)})(W(\beta_0)) + o_P(1);$
- (iii)  $\sqrt{n} \cdot \mathbf{P}_n \hat{B}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot q \cdot t_n = \sqrt{n} \cdot \mathbf{P}(\hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{T}_{\tilde{W}(\hat{\beta})}) \cdot \iota_t \cdot t_n + o_P(1);$
- (iv)  $\sqrt{n} \cdot \mathbf{P}_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot q \cdot t_n = \sqrt{n} \cdot \mathbf{P}(\hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{T}_{\hat{W}(\hat{\beta})}) \cdot \iota_t \cdot t_n + o_P(1);$
- (v)  $\sqrt{n} \cdot \mathbf{P}_n(t t_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot \iota_t \cdot f_{W(\beta_0)} \cdot t_n = -\sqrt{n} \cdot \mathbf{P}(\hat{T}_{\tilde{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\tilde{W}(\hat{\beta})}) \cdot \iota_t \cdot t_n - \sqrt{n} \cdot \mathbf{P}_n \partial t t_{W(\beta_0)} f_{W(\beta_0)} \iota_t \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1);$

The same results hold by replacing  $\tilde{W}(\hat{\beta})$  with  $\hat{W}(\hat{\beta})$ .

*Proof.* (i) By Lemma A.3, uniformly over  $\mathbf{T}$ ,

$$\begin{aligned} \mathbf{G}_{n\mathcal{E}} \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot q \cdot t_n &= \\ \mathbf{G}_{n\mathcal{E}} \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(W(\beta_0)) \cdot \varphi_t \cdot t_n &+ \mathbf{G}_{n\mathcal{E}} \partial f_{W(\beta_0)}(W(\beta_0)) \cdot \varphi_t \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) \cdot \iota_t + o_P(1). \end{aligned}$$

Conclude by applying Lemma A.5 to the first leading term of the above display, and Lemma A.9 to the second one.

(ii) By Lemma [A.3](#),

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot P_n \varepsilon t_n \int K(u) \cdot (f_{W(\beta_0)})(\tilde{W}(\hat{\beta}) + uh) du = \\ & - \frac{1}{\sqrt{n}} \cdot P_n \varepsilon \cdot t_n \int K(u) \cdot (f_{W(\beta_0)})(W(\beta_0) + uh) du + \\ & \frac{1}{\sqrt{n}} \cdot P_n \varepsilon \cdot \partial(f_{W(\beta_0)})(W(\beta_0)) \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1) \end{aligned}$$

uniformly over  $T$ . Conclude by applying Lemma [A.6](#) to the first leading term of the previous expression, and Lemma [A.9](#) to the second one..

(iii) By Lemma [A.3](#),

$$\begin{aligned} & - \frac{1}{\sqrt{n}} \cdot P_n \hat{B}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot t_n = \\ & \frac{1}{\sqrt{n}} \cdot P_n (\hat{T}_{\tilde{W}(\hat{\beta})} - t_{W(\beta_0)} \cdot \hat{T}_{\tilde{W}(\hat{\beta})})(W(\beta_0)) \cdot t_n + o_P(1) \end{aligned}$$

uniformly over  $T$ . Conclude by Lemma [A.5](#).

(iv) The proof is very similar to the one above.

(v) By Lemma [D.2](#), uniformly over  $T$

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot P_n (t_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot f_{W(\beta_0)} \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot P_n (\hat{T}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) - t_{W(\beta_0)}(W(\beta_0)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n + o_P(1). \end{aligned}$$

By Lemma [A.3](#) and Lemma [A.5](#) conclude for the result. □

**Lemma A.11.** Let  $H_0$  hold and assume that  $\frac{1}{\sqrt{n}} \cdot (\hat{\beta} - \beta_0) = o_P(1)$ . Under Assumption A1-A6, uniformly over  $T$

(i)

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})}) \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot P \varepsilon f_{W(\beta_0)} t_n + \frac{1}{\sqrt{n}} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - t_{W(\beta_0)} \hat{f}_{\tilde{W}(\hat{\beta})}) \cdot t_n + \\ & \frac{1}{\sqrt{n}} \cdot P_n \partial t_{W(\beta_0)} f_{W(\beta_0)} t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned}$$

(ii)

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})}) \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot P \varepsilon f_{W(\beta_0)} t_n + \frac{1}{\sqrt{n}} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - t_{W(\beta_0)} \hat{f}_{\tilde{W}(\hat{\beta})}) \cdot t_n + \\ & \frac{1}{\sqrt{n}} \cdot P_n \partial t_{W(\beta_0)} f_{W(\beta_0)} t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned}$$

*Proof.* By Assumption A5, Lemma [A.2](#), Lemma [D.3](#), Lemma [D.2](#), using the definitions of  $\hat{T}_{\tilde{W}(\hat{\beta})}$  and  $\hat{T}_{\tilde{W}(\hat{\beta})}$  and by the classical “change of variable”,

$$\frac{1}{\sqrt{n}} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})}) \cdot t_n = - \frac{1}{\sqrt{n}} \cdot P \int (Y - t_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n \cdot K(u) \cdot (f_{W(\beta_0)})(\tilde{W}(\hat{\beta}) + uh) du.$$

By adding and subtracting  $tt_{W(\beta_0)}(W(\beta_0))$ , and then using Lemma A.3, Lemma A.6, Lemma A.9, and Lemma D.1 uniformly over  $\mathcal{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n(Y - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (\iota_{tf_{W(\beta_0)}}(\tilde{W}(\hat{\beta}) + uh) du = \\ & \sqrt{n} \cdot P_n \varepsilon \iota_{tf_{W(\beta_0)}} + \\ & \sqrt{n} \cdot P_n(tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (\iota_{tf_{W(\beta_0)}}(W(\beta_0) + uh) du + o_P(1). \end{aligned}$$

Now, by the  $r$ -th order of the kernel  $K$ ,  $\int K(u) \cdot (\iota_{tf_{W(\beta_0)}}(w(\beta_0) + uh) du = (\iota_{tf_{W(\beta_0)}})(w(\beta_0)) + O(h^r)$ , with  $h^r n^{1/4} = o(1)$ , by Assumption A3. Hence, by Lemma D.2, uniformly over  $\mathcal{T}$

$$\begin{aligned} & \sqrt{n} \cdot P_n(tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (\iota_{tf_{W(\beta_0)}}(W(\beta_0) + uh) du = \\ & \sqrt{n} \cdot P_n(tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot t_n \cdot (\iota_{tf_{W(\beta_0)}})(W(\beta_0)) + o_P(1). \end{aligned}$$

Conclude by applying Lemma A.10 to the leading term of the previous expression. The proof of point (ii) proceeds along the same lines.  $\square$

In the next lemmas we prove the asymptotic expansion of  $\sqrt{n} P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \varphi_t \hat{t}_n$  and obtain the asymptotic behavior of the statistic.

**Lemma A.12.** *Let  $H_0$  hold and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = o_P(1)$ . Under Assumption A1-A5, uniformly over  $\mathcal{T}$*

$$\begin{aligned} & \sqrt{n} P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \varphi_t \hat{t}_n = \\ & \quad + \sqrt{n} \cdot P_n \varepsilon f_{V(\beta_0)} \varphi_t^\perp \\ & - \sqrt{n} \cdot P_n(\partial tt_{V(\beta_0)} f_{V(\beta_0)} \varphi_t^\perp) \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned}$$

*Proof.* First, by Lemma A.2 and D.2, we can replace the trimming  $\hat{t}_n$  with  $t_n$ . Then, by definition of  $\tilde{\varepsilon}$ ,

$$\begin{aligned} & \sqrt{n} P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \varphi_t \hat{t}_n = \\ & \sqrt{n} P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \varphi_t t_n + \sqrt{n} \cdot P_n(\hat{B}_{\tilde{W}(\hat{\beta})} \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot \varphi_t \cdot t_n. \end{aligned} \tag{A.9}$$

For the first term on the RHS, by adding and subtracting  $tt_{W(\beta_0)}(W(\beta_0))$  and using Lemma A.10, uniformly over  $\mathcal{T}$

$$\begin{aligned} & - \sqrt{n} P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \varphi_t t_n = \sqrt{n} \cdot P_n(Y - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \varphi_t \cdot t_n = \\ & \sqrt{n} P_n \varepsilon f_{W(\beta_0)} \varphi_t + \sqrt{n} \cdot P_n(tt_{W(\beta_0)}(W(\beta_0)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) - \hat{T}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\beta_0))) \cdot \varphi_t \cdot t_n + o_P(1). \end{aligned} \tag{A.10}$$

Applying Lemma A.3 to the second term on the RHS of the above display yields that, uniformly over  $\mathcal{T}$ ,

$$\begin{aligned} & - \sqrt{n} \cdot P_n(tt_{W(\beta_0)}(W(\beta_0)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) - \hat{T}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\beta_0))) \cdot \varphi_t \cdot t_n = \\ & \sqrt{n} \cdot P_n(tt_{W(\beta_0)} \cdot \hat{f}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n - \sqrt{n} \cdot P_n(\partial tt_{W(\beta_0)} f_{W(\beta_0)} \cdot \varphi_t) \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1) \end{aligned} \tag{A.11}$$

From Lemma A.5 the first term on the RHS of the above display can be approximated as

$$\sqrt{n} \cdot P_n(tt_W(\beta)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})} \cdot \varphi_t \cdot t_n = \sqrt{n} \cdot P(tt_W(\beta)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})} - \hat{T}_{\tilde{W}(\hat{\beta})} \cdot \varphi_t \cdot t_n, \quad (\text{A.12})$$

uniformly over  $T$ . Now, from Lemma A.10 and A.11, the bias correction term on the RHS of Eq. (A.9) can be approximated as

$$\begin{aligned} & - \sqrt{n} \cdot P_n(\hat{B}_{\tilde{W}(\hat{\beta})} \cdot \hat{f}_{\tilde{W}(\hat{\beta})})(\tilde{W}(\hat{\beta}))) \cdot \varphi_t \cdot t_n = \\ & - \sqrt{n} \cdot P(\varepsilon_{\tilde{W}(\hat{\beta})} f_{\tilde{W}(\hat{\beta})}) \cdot t_n + \sqrt{n} \cdot P(\hat{T}_{\tilde{W}(\hat{\beta})} - tt_W(\beta)) \cdot \hat{f}_{\tilde{W}(\hat{\beta})} \cdot t_n + \\ & \sqrt{n} \cdot P_n \partial tt_W(\beta_0) f_{W(\beta_0)} \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned} \quad (\text{A.13})$$

Plugging Eq. (A.13), (A.12), (A.11), and (A.10) into (A.9), and then rearranging terms yields the desired result.  $\square$

The following lemma provides the influence function representation for  $\sqrt{n} \cdot (\hat{\beta} - \beta_0)$ . Its proof is quite similar to those provided above, and is contained in the Supplementary Material.

**Lemma A.13.** *Under Assumptions A1-A6, and assuming that  $H_0$  holds,*

$$(i) \sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1);$$

(ii)

$$\begin{aligned} & - \sqrt{n} \cdot (\hat{\beta} - \beta_0) = \sqrt{n} \cdot P_n \varepsilon \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} tt_W(\beta_0) (W(\beta_0)) \\ & - \sqrt{n} \cdot P(D_2 - m(X)) \beta^T \cdot E\{\partial_1 tt_W(\beta) (W(\beta_0)) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} tt_W(\beta) (W(\beta))|_{\beta=\beta_0} | X_2\} + \\ & \sqrt{n} \cdot P_n(D_0 - m_0(Z)) \cdot E\{\partial_2 tt_W(\beta) (W(\beta_0)) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} tt_W(\beta) (W(\beta))|_{\beta=\beta_0} | Z\} + o_P(1), \end{aligned}$$

where  $\Sigma_0 := E\{\nabla_{\beta} tt_W(\beta_0)(W(\beta_0)) \cdot \nabla_{\beta}^T tt_W(\beta_0)(W(\beta_0))\}$ .

*Proof.* See Lemma ?? in the Supplementary Material.  $\square$

**Remark A.1.** In the following lines we will provide a more general version of Proposition 1. The expansion will refer to the model considered in this appendix, and will be based on the following function

$$\phi_{0,t}(y, \tilde{x}) := \sum_{y - tt_W(\beta_0)(w(\beta_0))}^{\Sigma} \cdot \psi_t(\tilde{x}) +$$

$$- E\{\psi_t(\tilde{X}) \cdot \partial_1 tt_W(\beta) (W(\beta_0)) | X_2 = x_2\} \cdot \beta_{0,2}^T \cdot (d_2 - m_2(x_2)) +$$

$$+ E\{\psi(X) \cdot \partial_2 tt_W(\beta) (W(\beta_0))^T \cdot Z = z\} \cdot (x^e - m_d(z)),$$

where  $\psi_t(\tilde{x}) := f_{W(\beta_0)}(w(\beta_0)) \cdot \varphi_t^\perp(\tilde{x}) - a(t)^T \Sigma_0^{-1} \nabla_{\beta} tt_W(\beta) (w(\beta))|_{\beta=\beta_0}$ ,  $\varphi_t(\tilde{x}) := \phi(\tilde{x} | t) - \iota_{W(\beta_0)}(w(\beta_0))$ ,  $\iota_{W(\beta_0)}(w) := E\{\phi(\tilde{X} | t) | W(\beta_0) = w\}$ ,  $\Sigma_0 := E\{\nabla_{\beta} tt_W(\beta)(w(\beta))|_{\beta=\beta_0} \cdot \nabla_{\beta}^T tt_W(\beta)(w(\beta))|_{\beta=\beta_0}\}$ , and finally



$a(t) := \int \partial_1 t \dot{W}(\beta_0) \cdot (w(\beta_0) - f_{W(\beta_0)}(w(\beta_0)) \cdot (x_1^T m_2(x_2))^T \varphi(x_1^T) dP(x)$ . The expansion in Proposition 1, as reported in the main text, can be obtained by imposing  $\beta_{0,2} = 0$  and recalling that  $\tilde{x} = (x, z)$ .

We are now able to provide a proof for Proposition 1.

### Proof of Proposition 1.

(i) From Lemma A.12, it is sufficient to obtain an expansion for  $\sqrt{n} \cdot P_n(\partial_1 t \dot{W}(\beta_0) f_{W(\beta_0)} \varphi_t^\perp) \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0))$ . To this end, notice that

$$\tilde{w}(\hat{\beta}) - w(\beta_0) = (x_1 \cdot (\hat{\beta}_1 - \beta_{0,1}) + \tilde{m}_2(x_2) \cdot (\hat{\beta}_2 - \beta_{0,2}) + \beta_{0,2} \cdot (\tilde{m}_2 - m_2)(x_2), (m_0 - \tilde{m}_0)(z)) \quad (A.14)$$

For the ease of notation, define

$$(\alpha_1(t), \alpha_2(t)) := (\partial_1 t \dot{W}(\beta_0) \cdot f_{W(\beta_0)} \varphi_t^\perp, \partial_2 t \dot{W}(\beta_0) \cdot f_{W(\beta_0)} \varphi_t^\perp) \quad (A.15)$$

The rates in Lemma D.1 and Lemma A.13 ensure that uniformly in  $t \in T$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n \partial_1 t \dot{W}(\beta_0) f_{W(\beta_0)} \varphi_t^\perp \cdot t_n \cdot (\tilde{W}(\hat{\beta}) - W(\beta_0)) = \\ & \sqrt{n} \cdot E\{(\alpha_1(t) \cdot X_1^T, \alpha_1(t) \cdot m_2(X_2)^T)\} \cdot \sqrt{n} \cdot (\hat{\beta} - \beta_0) + \\ & \sqrt{n} \cdot \beta_{0,2} \cdot P_n^T \alpha_1(t) \cdot (\tilde{m}_2 - m_2) - \sqrt{n} \cdot P_n \alpha_2(t) \cdot (\tilde{m}_0 - m_0) + o_p(1). \end{aligned}$$

By replacing the Influence Function Representations of Lemma A.13 and A.9 into the above display, and then using the expansion in Lemma A.12, the result is readily obtained.

(ii) Define

$$\begin{aligned} g_t(Y, D_0, D_2, \tilde{X}) &:= \varepsilon \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \varphi_t^\perp(\tilde{X}) \\ &- (D_2 - m_2(X_2)) \cdot \beta_{0,2} \cdot E\{\partial_1 t \dot{W}(\beta_0) \cdot (f_{W(\beta_0)} \varphi_t^\perp + a(t) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} \dot{W}(\beta_0)) \mid X_2\} + \\ & (D_0 - m_0(Z)) \cdot E\{\partial_2 t \dot{W}(\beta_0) \cdot (f_{W(\beta_0)} \varphi_t^\perp + a(t) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} \dot{W}(\beta_0)) \mid Z\}. \end{aligned}$$

To simplify, define  $\tilde{Z} := (Y, D_0, D_2, \tilde{X})$  and its support  $\tilde{Z} := Y \times D_0 \times D_2 \times \tilde{X}$ . Notice that under  $H_0$ ,  $P g_t = 0$  for all  $t \in T$ , so  $\sqrt{n} \cdot P_n g_t = G_n g_t$ . By the compactness of  $T$ , the continuity of  $\varphi$ , and the boundedness of the random variables involved,  $\sup_{\tilde{z} \in \tilde{Z}} |g_{t_1}(\tilde{z}) - g_{t_2}(\tilde{z})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in T$ . By Lemma (A.4) the class  $G := \{g_t : t \in T\}$  satisfies an entropy bound of the type  $\log N_{[]}(\delta, G, \|\cdot\|_\infty) \leq C \delta^{-v}$ , with  $v \in (0, 2)$  and the constants  $C$  and  $v$  which do not depend on  $\delta$ . Therefore, it follows by Donsker's Theorem (see Theorem 19.5 in van der Vaart (1998)),  $G_n \Rightarrow G$  over  $A^\infty(G)$ , where  $G$  is the Gaussian process defined by the Covariance Matrices collection  $\Psi(t_1, t_2) = \{P g_{t_1} g_{t_2} : t_1, t_2 \in T\}$ . Since the mapping  $f \mapsto \int |f \circ g_t|^2 \mu(dt)$  is continuous over  $A^\infty(G)$ ,

by a Continuous Mapping Theorem (see Theorem 18.11 in [van der Vaart \(1998\)](#)),

$$\int |G_n g_t|^2 \mu(dt) \rightarrow \int |G g_t|^2 \mu(dt) \quad (\text{A.16})$$

Similarly, since the operator  $\|\cdot\|_{G,\infty}$  is continuous over  $A^\infty(G)$ , by a Continuous Mapping Theorem,  $\|G_n\|_{G,\infty} \rightarrow \|G\|_{G,\infty}$ , with  $\|G\|_{G,\infty}$  tight process. Accordingly,  $\|G_n\|_{G,\infty} = o_P(1)$ . Define now

$$R_n(t) := \sqrt{n} P_n \tilde{e}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \phi_{\hat{t}_n} - \sqrt{n} P_n g_t.$$

From point (i),  $\|R_n\|_{T,\infty} = o_P(1)$ , so

$$\begin{aligned} S_n - \int |G_n g_t|^2 \mu(dt) &\leq \\ \int |R_n(t)| \cdot |R_n(t) + 2G_n g_t| \mu(dt) &= o_P(1). \end{aligned}$$

So,

$$S_n = \int |G_n g_t|^2 \mu(dt) + o_P(1) \quad (\text{A.17})$$

Conclude by Eq. (A.16), (A.17) and Slutsky's Theorem.

(iii) By Lemma C.2 and a reasoning similar to the one for Eq. (A.17),

$$\frac{S_n}{n} = \int P(Y - t t_{W(\beta^*)} (W(\beta^*))) \cdot f_{W(\beta^*)} (W(\beta^*)) \cdot \phi(t X) \mu^2(dt) + o_P(1)$$

By definition of  $H_1$ , the leading term on the RHS of the above display must be larger than zero.

[Q.E.D.]

## B Bootstrap Analysis

In the following lemmas, I consider an enlarged probability space which accounts for both the randomness of the original sample data and the bootstrap weights  $\zeta$ . For notational simplicity, let  $P^\zeta$  denote the product measure  $\otimes_{i=1}^n P^{\zeta_i}$ , and let  $P^\zeta \otimes P$  be the product measure between  $P^\zeta$  and the original measure  $P$ . Note the inconsistency of notation:  $P^\zeta$  here stands for the product measure  $\otimes_{i=1}^n P^{\zeta_i}$  instead of the probability measure of the single random element  $\zeta$ . We choose to accept such inconsistency as it enlightens the notational burdens.

Define

$$\begin{aligned} \tilde{W}(\beta) &:= (\beta \tilde{X} + \beta \tilde{m}(X_2^*), \tilde{X} - \tilde{m}(\tilde{Z})), & \hat{W}(\beta) &:= (\beta \tilde{X} + \beta \tilde{m}(X_2^*), \tilde{X} - \tilde{m}(\hat{Z})), \\ \tilde{w}(\beta) &:= (\beta \tilde{X} + \beta \tilde{m}(x_2^* - z), x - \tilde{m}_0^*(z)), & \text{and} \quad \hat{w}(\beta) &:= (\beta \tilde{X} + \beta \tilde{m}(x_2^*, x_2 - \hat{m}(z)))^* \end{aligned}$$

The lemmas that follow provide auxiliary results for the expansion of the empirical process at the basis of the

bootstrap version of the statistic.

**Lemma B.1.** Let  $H_0$  hold, and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$  and  $\sqrt{n} \cdot (\hat{\beta}^* - \beta^*) = O_P(1)$ . Under Assumptions A1-A6, uniformly over  $\{\tilde{x} : f(\tilde{x}) \geq \tau_n\}$

- (i)  $T_{\tilde{W}^*(\hat{\beta}^*)}^*(w^*(\hat{\beta}^*)) = T_{\tilde{W}^*(\hat{\beta}^*)}^*(w(\beta_0)) + \partial T_{\tilde{W}(\beta_0)}(w(\beta_0)) \cdot (w^*(\hat{\beta}^*) - w(\beta_0)) + o_P(n^{-1/2});$
- (ii)  $\hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(w^*(\hat{\beta}^*)) = \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(w(\beta_0)) + \partial f_{\tilde{W}(\beta_0)}(w(\beta_0)) \cdot (w^*(\hat{\beta}^*) - w(\beta_0)) + o_P(n^{-1/2});$
- (iii)  $\int K(u) \cdot (\iota f_{\tilde{W}(\beta_0)})(w^*(\hat{\beta}^*) + uh) du = \int K(u) \cdot (\iota f_{\tilde{W}(\beta_0)})(w(\beta_0) + uh) du + \partial(\iota f_{\tilde{W}(\beta_0)})(w(\beta_0)) \cdot (w^*(\hat{\beta}^*) - w(\beta_0)) + o_P(n^{-1/2}).$

*Proof.* The proof is very similar to the proof of Lemma (A.3) and is hence omitted.  $\square$

**Lemma B.2.** Let  $\zeta F := \{(\zeta, z) \mapsto \zeta \cdot f(z) : f \in F\}$ , where  $F$  is a class of functions. Then,  $N_{[\cdot]}(\|\zeta\|_{2,P} \cdot \delta, \zeta F, \|\cdot\|_{2,P}) \leq N_{[\cdot]}(\delta, F, \|\cdot\|_{2,P}).$

*Proof.* The proof proceeds along the same arguments as the proof of Lemma A.3 in [Escanciano et al. \(2014\)](#).  $\square$

**Lemma B.3.** Let  $H_0$  hold, and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$  and  $\sqrt{n} \cdot (\hat{\beta}^* - \beta^*) = O_P(1)$ . Under Assumptions A1-A6, uniformly in  $t \in T$ ,

- (i)  $G_{n,\zeta} \cdot t_n \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t = o_P(1);$
- (ii)  $G_{n,\zeta} \cdot \varepsilon \cdot t_n \cdot (f_{\tilde{W}^*(\hat{\beta}^*)}(W(\beta_0)) - f_{\tilde{W}(\beta_0)}(W(\beta_0))) \cdot \varphi_t \cdot t_n = o_P(1);$
- (iii)  $G_{n,\zeta} \cdot t_n \cdot (T_{\tilde{W}^*(\hat{\beta}^*)}^* - t t_{\tilde{W}(\beta_0)} \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}^*) \cdot \varphi_t = o_P(1);$
- (iv)  $\sqrt{n} \cdot P_{n,\zeta} \cdot \varepsilon \cdot \int K(u) \cdot (\iota f_{\tilde{W}(\beta_0)})(W(\beta_0) + uh) du = \sqrt{n} \cdot P_{n,\zeta} \cdot \varepsilon \cdot \iota f_{\tilde{W}(\beta_0)} + o_P(1);$
- (v)  $\sqrt{n} P_{n,\zeta} \cdot (D_0 - m_0) \cdot \int K_0(u) \cdot \phi_t(Z + uh_0) du = \sqrt{n} P_{n,\zeta} \cdot (D_0 - m_0) \cdot \phi_t + o_P(1).$

*Proof.* (i) Let  $C_n$  denote the class  $C(W_n)$ . By Lemma D.3 and Tonelli-Fubini's Theorem,  $\|\zeta \cdot t \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t\|_{2,P} = \|\zeta\|_{2,P} \cdot \|t_n \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t\|_{2,P} \leq \|\zeta\|_{2,P} \cdot o_P(1)$ . Hence, by Lemma D.4, for any fixed  $\delta > 0$ , the event

$$B_n := \{\sup_{t \in T} \|\zeta \cdot t_n \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t\|_{2,P} < \delta, t_n \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \in t_n \cdot C_n\}$$

has a probability converging to one as  $n \rightarrow \infty$ . Notice that whenever  $B_n$  holds,

$$\sup_{t \in T} |G_{n,\zeta} \cdot t_n \cdot (T_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t| \leq \|G_n\|_{(\zeta \cdot t_n \cdot C_n \Phi)^\delta},$$

with  $(\zeta \cdot t_n \cdot C_n \cdot \Phi)^\delta$  defined similarly as in Lemma A.5. Since  $(\zeta \cdot t_n \cdot C_n \cdot \Phi)^\delta \subset \zeta \cdot (t_n \cdot C_n \cdot \Phi)$ ,

$$N_{[\cdot]}(\delta, (\zeta \cdot t_n \cdot C_n \cdot \Phi)^\delta, \|\cdot\|_{2,P}) \leq N_{[\cdot]}(\delta, \zeta \cdot t_n \cdot C_n \cdot \Phi, \|\cdot\|_{2,P}).$$

By the proof of Lemma A.5,  $N(\delta, t_n C_n, \|\cdot\|_\infty) \leq N(\delta, C_n, \|\cdot\|_{W_n, \infty})$ . So, from Lemma B.2, Eq. (A.7), and Lemma A.4,  $\log N_{[\cdot]}(\delta, \zeta \cdot t_n \cdot C_n \cdot \Phi, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v}$ , with  $C$  and  $v$  being constants which do not depend on  $\delta$ , and  $v \in (0, 2)$ . From this point proceed as in Lemma A.5 to obtain point (i). The proofs of Point (ii) and (iii)

proceed in essentially the same way.

(iv) Define  $g_{n,t}(\tilde{x}, \varepsilon) := \varepsilon \cdot \int K(u) \cdot (\iota_{f_W(\beta_0)})(w(\beta_0) + uh) du - \varepsilon \cdot (\iota_{f_W(\beta_0)})(w(\beta_0))$ . From the proof of Lemma A.6,  $\sup_{t \in T} \|g_{n,t}\|_{2,P} = o(1)$ . Hence, by Tonelli-Fubini's Theorem,  $\sup_{t \in T} \|\zeta \cdot g_{n,t}\|_{2,P} = \|\zeta\|_{2,P} \cdot \sup_{t \in T} \|g_{n,t}\|_{2,P} = o(1)$ . From the proof of Lemma (A.6),  $\log N_{[\cdot]}(\delta, G_n, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v}$ , with  $G_n := \{g_{n,t} : t \in T\}$ ,  $C$  and  $v$  being constants which do not depend on neither  $n$  nor  $\delta$ . Hence, from Lemma B.2,  $\log N_{[\cdot]}(\delta, \zeta \cdot G_n, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v}$  (by changing the constant  $C$  accordingly). From this point, conclude by proceeding in as in Lemma A.5.

(v) The proof is identical to the one of Point (iv), so it is omitted.  $\square$

**Lemma B.4.** Let  $(\tilde{x}, t) \mapsto \phi_t(\tilde{x})$  be a fixed function with  $\sup_{t \in T, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ , such that  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in T$ . Under Assumptions A1-A6,

(i)  $G_n(\hat{m}_0^* - m_0) \cdot t_n \cdot \phi_t = o_P(1)$ ;

(ii)  $G_n \zeta \cdot (\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t = o_P(1)$ .

The same result holds for  $\hat{m}_2^*$  and  $\hat{m}_2$ .

*Proof.* (i) The proof is identical to the proof of Lemma (A.7)

(ii) Let  $C_n$  denote  $C(Z_n)$ , and let  $Y := \{\phi_t : t \in T\}$ . By the same reasoning as in the proof of Lemma (A.7),  $\log N_{[\cdot]}(\delta, t_n \cdot (C_n - m_0) \cdot Y, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v}$ . So Lemma (B.2) ensures that

$$\log N_{[\cdot]}(\delta, \zeta \cdot t_n \cdot (C_n - m_0) \cdot Y, \|\cdot\|_{2,P}) \leq C \cdot \delta^{-v} \text{ with } v \in (0, 2).$$

From Lemma D.1, for all  $t \in T$  the mapping  $(\zeta, \tilde{x}) \mapsto \zeta \cdot (\hat{m}_0 - m_0)(z) \cdot t_n(\tilde{x}) \cdot \phi_t(\tilde{x})$  belongs to the class  $\zeta \cdot t_n \cdot (C_n - m_0) \cdot Y$  with a probability converging to one. For the  $L_2$ -convergence, from Lemma D.1, the boundedness of  $\phi_t$  and Tonelli-Fubini's Theorem,

$$\sup_{t \in T} \|\zeta \cdot (\hat{m}_0 - m_0) \cdot \phi_t\|_{2,P} \leq \|\zeta\|_{2,P} \cdot o_P(1).$$

So, by following the same steps as in the proof of Lemma A.5, the desired result is obtained.  $\square$

**Lemma B.5.** Let  $(\tilde{x}, t) \mapsto \phi_t(\tilde{x})$  be a fixed function with  $\sup_{t \in T, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ , such that  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in T$ . Under Assumptions A1-A6,

$$\begin{aligned} & \sqrt{n} \cdot P_n \phi_t \cdot (\hat{m}_0^* - \hat{m}_0) \cdot t_n = \\ & \sqrt{n} \cdot P_n (\hat{m}_0 - m_0) \cdot t_n \cdot E\{\phi_t(\tilde{X}) | Z\} + \\ & \sqrt{n} \cdot P_n \zeta \cdot (D_0 - m_0) \cdot E\{\phi_t(\tilde{X}) | Z\} \\ & - \sqrt{n} \cdot P_n (D_0 - m_0) \cdot E\{\phi_t(\tilde{X}) | Z\} + o_P(1) \end{aligned}$$

uniformly in  $t \in \mathbb{T}$ .

*Proof.* By adding and subtracting  $m_0$ , and then using Lemma B.4, Lemma A.7, uniformly over  $\mathbb{T}$

$$\sqrt{n} \cdot P_n(\hat{m}_0^* - \hat{m}_0) \cdot t_n = o_P(1) + \sqrt{n} \cdot P(\hat{m}_0^* - m_0) \cdot t_{n,Z}^{(\eta/2)} \cdot \phi_t.$$

For simplicity, denote  $\phi_t^{\sim}(Z) := E\{\phi_t(X^{\sim}) \mid Z\}$ . From Lemma D.1, and Lemma A.2, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P(\hat{m}_0^* - \hat{m}_0) \cdot t_n \cdot \phi_t = \\ & \sqrt{n} \cdot \int (\hat{T}_0^* - \hat{T}_0)(z) \cdot t_{n,Z}^{(\eta/2)}(z) \cdot \phi_t^{\sim}(z) dz + o_P(1). \end{aligned} \quad (\text{B.18})$$

By Lemma A.2, definition of  $\hat{T}_0^*$  and  $\hat{T}_0$  and the usual “change of variable”, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} \sqrt{n} \cdot \int (\hat{T}_0^* - \hat{T}_0)(z) \cdot t_{n,Z}^{(\eta/2)}(z) \cdot \phi_t^{\sim}(z) dz &= \sqrt{n} \cdot P_n \xi \cdot (D_0 - m_0) \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du + \\ &- \sqrt{n} \cdot P_n \xi \cdot (\hat{m}_0 - m_0) \cdot t_n \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du + \\ &\sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du \\ &- \sqrt{n} P_n(D_0 - m_0) \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du. \end{aligned}$$

Lemma D.1, the  $r_0$ -th order of the kernel, and the usual  $r_0$ -th order Taylor expansion yield

$$\begin{aligned} \sqrt{n} \cdot P_n \xi \cdot (\hat{m}_0 - m_0) \cdot t_n \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du &= \\ \sqrt{n} \cdot P_n \xi \cdot (\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t^{\sim}(Z) + o_P(1), \end{aligned}$$

uniformly over  $\mathbb{T}$ . Similarly, uniformly over  $\mathbb{T}$

$$\begin{aligned} \sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \int K(u) \cdot \phi_t^{\sim}(Z + uh) du &= \\ \sqrt{n} \cdot P_n(\hat{m}_0 - m_0) \cdot t_n \cdot \phi_t^{\sim}(Z) + o_P(1) \end{aligned}$$

Finally, by Lemma B.3 and A.6, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} \sqrt{n} \cdot P_n \xi \cdot (D_0 - m_0) \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du &= \\ \sqrt{n} \cdot P_n \xi \cdot (D_0 - m_0) \cdot \phi_t^{\sim}(Z) + o_P(1). \end{aligned}$$

$$\begin{aligned} \sqrt{n} \cdot P_n(D_0 - m_0) \cdot \int K_0(u) \cdot \phi_t^{\sim}(Z + uh_0) du &= \\ \sqrt{n} \cdot P_n(D_0 - m_0) \cdot \phi_t^{\sim}(Z) + o_P(1). \end{aligned}$$

Conclude by the last five displays. □

**Lemma B.6.** Let  $(\tilde{x}, t) \mapsto \phi_t(\tilde{x})$  be a fixed function with  $\sup_{t \in \mathbb{T}, \tilde{x} \in \text{Supp}(\tilde{X})} |\phi_t(\tilde{x})| < \infty$ , such that  $\sup_{\tilde{x} \in \text{Supp}(\tilde{X})} |\phi_{t_1}(\tilde{x}) - \phi_{t_2}(\tilde{x})| \leq C \cdot |t_1 - t_2|$  for all  $t_1, t_2 \in \mathbb{T}$ . Under Assumptions A1-A6, uniformly over  $\mathbb{T}$

$$\sqrt{n} \cdot \mathbb{P}_n \phi_t \cdot t_n \cdot (\tilde{m}_0^* - \hat{m}_0) = \sqrt{n} \cdot \mathbb{P}_n \xi \cdot (D_0 - m_0) \cdot \mathbb{E}\{\phi_t(\tilde{X}) \mid Z\} + o_P(1).$$

The same result holds for  $\tilde{m}_2^*$

*Proof.* From the definition of  $\tilde{m}_2^*$

$$\sqrt{n} \cdot \mathbb{P}_n \phi_t \cdot t_n \cdot (\tilde{m}_0^* - \hat{m}_0) = \sqrt{n} \cdot \mathbb{P}_n \phi_t \cdot t_n \cdot (\hat{m}_0^* - \hat{m}_0) - \sqrt{n} \cdot \mathbb{P}_n \phi_t \cdot t_n \cdot \hat{B}_0$$

Conclude from Lemma (B.5) and Lemma (A.8). □

**Lemma B.7.** Let  $H_0$  hold, and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$  and  $\sqrt{n} \cdot (\hat{\beta}^* - \beta^*) = O_P(1)$ . Under Assumptions A1-A6, uniformly over  $\mathbb{T}$ ,

- (i)  $\sqrt{n} \cdot \mathbb{P}_n \xi \cdot \varepsilon \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t = \sqrt{n} \cdot \mathbb{P}_n \xi \cdot \varepsilon \cdot f_{W(\beta_0)} \cdot \varphi_t + o_P(1);$
- (ii)  $\sqrt{n} \cdot \mathbb{P}_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t = o_P(1);$
- (iii)  $\sqrt{n} \cdot \mathbb{P}_n (\hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) - tt_{W(\beta_0)}(W(\beta_0))) \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t =$   
 $\sqrt{n} \mathbb{P}(\hat{T}_{\hat{W}(\hat{\beta})} - tt_{W(\beta_0)} \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}) \varphi_t t_n + \sqrt{n} \mathbb{P}_n \partial tt_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (W(\hat{\beta}) - W(\beta_0)) + o_P(1);$
- (iv)  $\sqrt{n} \cdot \mathbb{P}_n \xi \cdot \varepsilon \cdot t_n \cdot \int K(u) \cdot (tt_{W(\beta_0)})(\tilde{W}^*(\hat{\beta}^*) + uh) du = \sqrt{n} \cdot \mathbb{P}_n \xi \cdot \varepsilon \cdot t_n \cdot f_{W(\beta_0)} + o_P(1);$
- (v)  $\sqrt{n} \cdot \mathbb{P}_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (f_{W(\beta_0)} t_t)(\tilde{W}^*(\hat{\beta}^*) + uh) du = o_P(1).$
- (vi)  $\sqrt{n} \cdot \mathbb{P}_n (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}^*(\tilde{W}^*(\hat{\beta}^*))) \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t =$   
 $\sqrt{n} \cdot \mathbb{P} (tt_{W(\beta_0)} \hat{f}_{\tilde{W}^*(\hat{\beta}^*)} - \hat{T}_{\tilde{W}^*(\hat{\beta}^*)}) \cdot t_n \cdot \varphi_t - \sqrt{n} \cdot \mathbb{P}_n \partial tt_{W(\beta_0)} f_{W(\beta_0)} \varphi_t \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + o_P(1).$

*Proof.* (i) From Lemma B.1,

$$\sqrt{n} \mathbb{P}_n \xi \cdot \varepsilon \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t = \mathbb{G} \xi \cdot \varepsilon \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(W(\beta_0)) \cdot t_n \cdot \varphi_t +$$

$$\sqrt{n} \cdot \mathbb{P}_n \xi \cdot \varepsilon \cdot \partial f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + R_{n,t},$$

with  $\sup_{t \in \mathbb{T}} |R_{n,t}| \leq o_P(n^{-1/2})$ .  $\sqrt{n} \cdot \mathbb{P}_n |\xi| = o_P(1)$ . By applying Lemma B.6 and Lemma A.7, the second term on the RHS of the above display is  $o_P(1)$ . Conclude by applying Lemma to the first term on the RHS.

(ii) From Lemma D.2, uniformly over the set  $\mathbb{T}$

$$\sqrt{n} \cdot \mathbb{P}_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\tilde{W}^*(\hat{\beta}^*)}(\tilde{W}^*(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t =$$

$$\sqrt{n} \cdot \mathbb{P}_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \varphi_t + o_P(1).$$

By Lemma B.1,

$$\begin{aligned} & \sqrt{n} \cdot P_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \varphi_t = \\ & \quad - \sqrt{n} \cdot P_n \xi \cdot (tt_{W(\beta_0)} \cdot \hat{f}_{\hat{W}(\hat{\beta})} - \hat{T}_{\hat{W}(\hat{\beta})}) \cdot t_n \cdot \varphi_t \\ & \quad - \sqrt{n} \cdot P_n \xi \cdot \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned}$$

Conclude by applying Lemma B.3 and Lemma A.7 to the two leading terms of the latter expansion.

(iii) The proof follows the same steps as the proof of point (ii).

(iv) By Lemma B.1, uniformly over  $T$

$$\begin{aligned} & \sqrt{n} \cdot P_n \xi \cdot \varepsilon \cdot t_n \cdot \int K(u) \cdot (\iota_{tf_{W(\beta_0)}})(\tilde{W}^*(\hat{\beta}^*) + uh) du = \\ & \quad \sqrt{n} \cdot P_n \xi \cdot \varepsilon \cdot t_n \cdot \int K(u) \cdot (\iota_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du + \\ & \quad \sqrt{n} \cdot P_n \xi \cdot \varepsilon \cdot \partial(\iota_{tf_{W(\beta_0)}})(W(\beta_0)) \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + o_P(1). \end{aligned}$$

By applying Lemma B.6 and Lemma A.7, the second term on the RHS of the above display is  $o_P(1)$ . Conclude by applying Lemma B.3 to the first leading term of the latter expansion.

(v) By Lemma B.1 and D.2,

$$\begin{aligned} & -\sqrt{n} \cdot P_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (f_{W(\beta_0)\iota_t})(\tilde{W}^*(\hat{\beta}^*) + uh) du = \\ & \quad \sqrt{n} \cdot P_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (f_{W(\beta_0)\iota_t})(W(\beta_0) + uh) du + o_P(1) = \\ & \quad \sqrt{n} \cdot P_n \xi \cdot (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot (f_{W(\beta_0)\iota_t})(W(\beta_0) + uh) + o_P(1) \end{aligned}$$

uniformly in  $t \in T$ , where the last equality follows from an  $r$ -th order Taylor expansion, the  $r$ -th order of the kernel, Assumption A4, and Lemma D.2. By proceeding similarly to the proof of Point (ii), we conclude.

(vi) By Lemma B.1, uniformly over  $T$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n (tt_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}^*(\tilde{W}^*(\hat{\beta}^*))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}^*(\tilde{W}^*(\hat{\beta}^*))) \cdot t_n \cdot \varphi_t = \\ & \quad \sqrt{n} \cdot P_n (tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}^* - \hat{T}_{\hat{W}(\hat{\beta})}^*) \cdot t_n \cdot \varphi_t \\ & \quad - \sqrt{n} \cdot P_n \partial tt_{W(\beta_0)} f_{W(\beta_0)} \varphi_t \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + o_P(1). \end{aligned}$$

An application of Lemma B.3 to the first leading term yields the desired result □

**Lemma B.8.** Under Assumption A1-A6,  $H_0$ ,  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$ , and  $\sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) = O_P(1)$ , uniformly over  $T$ ,

(i)

$$\begin{aligned}
& \sqrt{n} \cdot P(T_{\tilde{W}(\hat{\beta}^*)}^* - tt_{W(\beta_0)} f_{\tilde{W}(\hat{\beta}^*)}^*) \cdot t_n \cdot l_t = \\
& \sqrt{n} \cdot P_{\hat{W}(\hat{\beta})} \int K(u) \cdot (l_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du - \\
& \sqrt{n} \cdot P_n \int K(u) \cdot (tt_{W(\beta_0)} l_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du + \\
& - \sqrt{n} \cdot P_n \partial tt_{W(\beta_0)} \cdot l_t \cdot f_{W(\beta_0)} \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + \sqrt{n} \cdot P_n \xi \varepsilon l_{tf_{W(\beta_0)}} + o_P(1)
\end{aligned}$$

(ii)

$$\begin{aligned}
& \sqrt{n} P(T_{\hat{W}(\hat{\beta})} - tt_{W(\beta_0)} f_{\hat{W}(\hat{\beta})}) l_t \cdot t_n = \\
& \sqrt{n} \cdot P_n Y \int K(u) \cdot (l_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du - \sqrt{n} \cdot P_n \int K(u) \cdot (tt_{W(\beta_0)} f_{W(\beta_0)} l_t)(W(\beta_0) + uh) du \\
& - \sqrt{n} \cdot P_n \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot l_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1)
\end{aligned}$$

Proof. (i) By Lemma A.2, the definition of  $T_{\tilde{W}(\hat{\beta}^*)}^*$  and  $\hat{f}_{\tilde{W}(\hat{\beta}^*)}$ , and the classical change of variable,

$$\begin{aligned}
& \sqrt{n} \cdot P(T_{\tilde{W}(\hat{\beta}^*)}^* - tt_{W(\beta_0)} f_{\tilde{W}(\hat{\beta}^*)}^*) \cdot t_n \cdot l_t = \\
& - \sqrt{n} \cdot P_{\hat{W}(\hat{\beta})} \int K(u) \cdot (l_{tf_{W(\beta_0)}})(\tilde{W}^*(\hat{\beta}^*) + uh) du + \\
& \sqrt{n} \cdot P_n \xi \cdot (Y - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (l_{tf_{W(\beta_0)}})(\tilde{W}^*(\hat{\beta}^*) + uh) du + \\
& - \sqrt{n} \cdot P_n t_n \cdot \int K(u) \cdot (tt_{W(\beta_0)} l_{tf_{W(\beta_0)}})(\tilde{W}^*(\hat{\beta}^*) + uh) du =: A_{n,1} + A_{n,2} - A_{n,3}.
\end{aligned} \tag{B.19}$$

By Lemma B.1, Lemma D.1, and Lemma D.2, uniformly in  $t \in T$ ,

$$\begin{aligned}
A_{n,1} &= \sqrt{n} \cdot P_{\hat{W}(\hat{\beta})} \int K(u) \cdot (l_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du + \\
& \sqrt{n} \cdot tt_{W(\beta_0)}(W(\beta_0)) \cdot \partial(l_{tf_{W(\beta_0)}})(W(\beta_0)) \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + o_P(1).
\end{aligned} \tag{B.20}$$

Similarly, by Lemma B.1, uniformly over  $T$ ,

$$\begin{aligned}
A_{n,3} &= \sqrt{n} \cdot P_n \int K(u) \cdot (tt_{W(\beta_0)} l_{tf_{W(\beta_0)}})(W(\beta_0) + uh) du + \\
& + \sqrt{n} \cdot P_n \partial(tt_{W(\beta_0)} l_{tf_{W(\beta_0)}}) \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - W(\beta_0)) + o_P(1).
\end{aligned} \tag{B.21}$$

Finally, by Lemma B.7, uniformly in  $t \in T$ ,

$$A_{n,2} = \sqrt{n} \cdot P_n \xi \varepsilon l_{tf_{W(\beta_0)}} + o_P(1). \tag{B.22}$$

Conclude for Point (i) by plugging Eq. (B.22), (B.21), and (B.20) into (B.19).



(ii) By definition of  $\hat{T}_{\hat{W}(\hat{\beta})}$  and  $\hat{f}_{\hat{W}(\hat{\beta})}$ , the classical change of variable and Lemma A.2,

$$\begin{aligned} & - \sqrt{n} P \left( \hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})} \right) \cdot t_n \cdot \iota_t = \\ & \sqrt{n} \cdot P_n Y \cdot t_n \cdot \int K(u) \cdot (\iota_t f_{W(\beta_0)})(\hat{W}(\hat{\beta}) + uh) du + \\ & - \sqrt{n} P_n t_n \cdot \int (t t_{W(\beta_0)} f_{W(\beta_0)} \iota_t)(\hat{W}(\hat{\beta}) + uh) du. \end{aligned}$$

By applying Lemma B.1 to each of the two term on the RHS of the above expression, and then rearranging,

$$\begin{aligned} & - \sqrt{n} P \left( \hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})} \right) \cdot t_n \cdot \iota_t = \\ & \sqrt{n} \cdot P_n Y \cdot \int K(u) \cdot (\iota_t f_{W(\beta_0)})(W(\beta_0) + uh) du + \\ & \sqrt{n} \cdot P_n \varepsilon \partial(\iota_t f_{W(\beta_0)}) \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + \\ & - \sqrt{n} \cdot P_n \int (t t_{W(\beta_0)} f_{W(\beta_0)} \iota_t)(W(\beta_0) + uh) du \\ & - \sqrt{n} \cdot P_n \partial t t_{W(\beta_0)} \cdot \iota_t \cdot f_{W(\beta_0)} \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1) \end{aligned}$$

uniformly over  $\mathcal{T}$ . By using  $H_0$ , Lemma A.13, and Lemma A.7 it is immediate to show that the second term on the RHS of the above display is  $o_P(1)$  uniformly over  $\mathcal{T}$ . Hence, point (ii) follows.  $\square$

**Lemma B.9.** Let  $H_0$  hold and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$  and  $\sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) = O_P(1)$ . Under Assumptions A1-A6, uniformly over  $\mathcal{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n \varepsilon^* \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_{it} \cdot \hat{\pi} \\ & \sqrt{n} \cdot P_n \varepsilon \varphi^\perp_t + \\ & - \sqrt{n} \cdot P_n \partial t t_{W(\beta_0)} f_{W(\beta_0)} \iota_t \cdot t_n \cdot (\tilde{W}(\hat{\beta}^*) - \hat{W}(\hat{\beta})) + o_P(1). \end{aligned}$$

*Proof.* By Lemma A.2 and D.2, we can replace the trimming  $\hat{t}_n$  with  $t_n$ . So, we obtain the following decomposition

$$\begin{aligned} & - \sqrt{n} \cdot P_n \varepsilon^* \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_{it} \cdot \hat{\pi} \\ & \sqrt{n} \cdot P_n (Y^* - t_n^* \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*))) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*)) \cdot t_n \cdot \varphi_t + \\ & + \sqrt{n} \cdot P_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_{it} \cdot t_n \end{aligned} \quad (B.23)$$

By the rates in Lemma D.2 uniformly in  $t \in \mathcal{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_{it} \cdot \hat{\pi} \\ & \sqrt{n} \cdot P_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \varphi_{it} \cdot t_n + o_P(1). \end{aligned} \quad (B.24)$$

By Lemma A.10 and Lemma A.11, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \varphi_t \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \hat{f}_{W(\beta_0)} \cdot \varphi_t + \frac{1}{\sqrt{n}} \cdot \mathbf{P}(T_{\hat{W}(\hat{\beta})} - tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n + \\ & + \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned} \quad (\text{B.25})$$

Now, the first term on the RHS of Eq. (B.23) can be decomposed as

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(Y - t_{\hat{W}(\hat{\beta})}^*(\tilde{W}(\hat{\beta}^*))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \xi \cdot (Y - t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \cdot t_n + \\ & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) - t_{\hat{W}(\hat{\beta})}^*(\tilde{W}(\hat{\beta}^*))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \cdot t_n \end{aligned} \quad (\text{B.26})$$

By Lemma B.7, uniformly in  $t \in \mathbb{T}$ ,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \xi \cdot (Y - t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \cdot t_n = \\ & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \xi \varepsilon f_{W(\beta_0)} \varphi_t + o_P(1). \end{aligned} \quad (\text{B.27})$$

Also, Lemma B.7 implies that, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) - t_{\hat{W}(\hat{\beta})}^*(\tilde{W}(\hat{\beta}^*))) \cdot \hat{f}_{\hat{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \cdot t_n = \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}(T_{\hat{W}(\hat{\beta})} - tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n + \\ & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}(T_{\hat{W}(\hat{\beta})}^* - tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\tilde{W}(\hat{\beta}^*) - \hat{W}(\hat{\beta})) + o_P(1). \end{aligned} \quad (\text{B.28})$$

Now, the two bias integrals in the above display can be handled by using Lemma B.8. Thus, uniformly in  $t \in \mathbb{T}$ ,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \mathbf{P}(T_{\hat{W}(\hat{\beta})} - tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n - \frac{1}{\sqrt{n}} \cdot \mathbf{P}(T_{\hat{W}(\hat{\beta})}^* - tt_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot \varphi_t \cdot t_n = \\ & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(Y - t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (\iota_t f_{W(\beta_0)})(W(\beta_0) + uh) du \\ & - \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \xi \varepsilon \iota_t f_{W(\beta_0)} + \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \partial tt_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t \cdot t_n \cdot (\tilde{W}(\hat{\beta}^*) - \hat{W}(\hat{\beta})) + o_P(1). \end{aligned} \quad (\text{B.29})$$

By adding and subtracting  $tt_{W(\beta_0)}(W(\beta_0))$ , and then using Lemma B.3, Lemma D.2, the  $r$ -th order of the kernel  $K$ , and Assumption A5,

$$\begin{aligned} & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(Y - t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot t_n \cdot \int K(u) \cdot (\iota_t f_{W(\beta_0)})(W(\beta_0) + uh) du = \\ & \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n \varepsilon f_{W(\beta_0)} \varphi_t + \frac{1}{\sqrt{n}} \cdot \mathbf{P}_n(tt_{W(\beta_0)}(W(\beta_0)) - t_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot \varphi_t \cdot f_{W(\beta_0)} \cdot t_n + o_P(1). \end{aligned}$$

By Lemma D.2, we can replace  $f_{W(\beta_0)}(W(\beta_0))$  with  $\hat{f}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))$  in the second term on the RHS of the above

display. Then, by using Lemma B.1 and Lemma B.3, uniformly in  $t \in \mathbb{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n(t t_{W(\beta_0)}(W(\beta_0)) - \hat{t}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta}))) \cdot l_t \cdot f_{W(\beta_0)} \cdot t_n = \\ & \quad -P(\hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot l_t \cdot t_n \\ & - \sqrt{n} \cdot P_n \partial t t_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot l_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned} \quad (B.30)$$

Putting together B.26, B.27, B.28, B.29, and B.30, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} & \sqrt{n} \cdot P_n(Y \stackrel{*}{=} t \hat{W}^*(\hat{\beta}^*)) \cdot f_{\hat{W}^*(\hat{\beta}^*)}(\hat{W}^*(\hat{\beta}^*)) \cdot t \cdot \varphi = \\ & \sqrt{n} \cdot P_n \xi \varepsilon f_{W(\beta_0)} \varphi^\perp_t - \sqrt{n} \cdot P_n \partial t t_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot \varphi_t^\perp \cdot t_n \cdot (\hat{W}^*(\hat{\beta}^*) - \hat{W}(\hat{\beta})) + \\ & \sqrt{n} \cdot P_n \xi \hat{f}_{W(\beta_0)} - \sqrt{n} \cdot P(\hat{T}_{\hat{W}(\hat{\beta})} - t t_{W(\beta_0)} \hat{f}_{\hat{W}(\hat{\beta})}) \cdot l_t \cdot t_n + \\ & - \sqrt{n} \cdot P_n \partial t t_{W(\beta_0)} \cdot f_{W(\beta_0)} \cdot l_t \cdot t_n \cdot (\hat{W}(\hat{\beta}) - W(\beta_0)) + o_P(1). \end{aligned} \quad (B.31)$$

Plugging Eq. (B.31), (B.24), and (B.25) into (B.23) yield the desired result.  $\square$

The following Lemma provides the Influence Function Representation for  $\sqrt{n} \cdot (\hat{\beta}^* - \beta_0)$ . Its proof is similar to the proofs of the lemmas provided above, so it is reported in a supplementary material.

**Lemma B.10.** *Let  $H_0$  hold, and assume that  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$ . Under Assumptions A1-A6,*

$$\begin{aligned} & - \\ & \sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) = \sqrt{n} \cdot P_n \xi \cdot \varepsilon \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} t_{\hat{W}(\beta)} \\ & - \sqrt{n} \cdot P_n \xi \cdot (D_{\mathbb{Z}} m(\mathbb{Z})) \beta^T \cdot E\{\partial_1 t_{\hat{W}(\beta)}(W(\beta_0)) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} t_{\hat{W}(\beta)}(W(\beta))\}_{\beta=\beta_0} | X_2\} + \\ & \sqrt{n} \cdot P_n \xi \cdot (D_0 - m_0(Z)) \cdot E\{\partial_2 t_{W(\beta_0)}(W(\beta_0)) \cdot \Sigma_0^{-1} \cdot \nabla_{\beta} t_{\hat{W}(\beta)}(W(\beta))\}_{\beta=\beta_0} | Z\} + o_P(1). \end{aligned}$$

*Proof.* See Lemma ?? in the Supplementary Material.  $\square$

**Lemma B.11.** *Define  $g_t$  and  $\mathbf{G}$  as in the proof of Proposition 1. Let  $\mathbf{G} := \{g_t : t \in \mathbb{T}\}$ . Under Assumption A1,  $\mathbf{G}_n^{(\xi)} : \mathbf{G}$  uniformly over  $l(\mathbf{G})$  wrt the joint measure  $\mathbf{P} \otimes P$ .*

*Proof.* The proof follows the same steps as the proof of Proposition 1 Point (ii). The only difference is that for the derivation of the entropy condition both Lemma A.4 and Lemma B.2 are used.  $\square$

**Lemma B.12.** *Let  $\hat{Z}$  and  $Z$  be two random variables with respect to  $\mathbf{P}^\xi \otimes P$  that take value in  $\mathbf{R}$ . Let  $Q$  be a probability measure defined over the Borel sigma field generated by  $\mathbf{R}$ . Denote with  $L_Q$  the cdf generated by the measure  $Q$ . If*

- (i)  $\mathbf{P}_n^\xi(|\hat{Z} - Z| > \delta) \rightarrow^P 0$  for all  $\delta > 0$ ,
- (ii)  $\mathbf{P}_n^\xi(Z \leq z) \rightarrow^P L_Q(z)$  for all  $z$ ,
- (iii)  $L_Q(\cdot)$  is a continuous function,

then

$$\sup_{z \in \mathbb{R}} |\mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) - L_Q(z)| = o_P(1).$$

*Proof.* For a fixed  $\delta > 0$ ,

$$\begin{aligned} \mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) &\leq \mathbb{P}_n^{\tilde{\zeta}}(\{Z \leq z + Z - \hat{Z}\} \cap \{|\hat{Z} - Z| \leq \delta\}) + \\ &\quad \mathbb{P}_n^{\tilde{\zeta}}(\{|\hat{Z} - Z| > \delta\}) \leq \mathbb{P}_n^{\tilde{\zeta}}(Z \leq z + \delta) + o_P(1). \end{aligned}$$

By a similar reasoning,

$$\mathbb{P}_n^{\tilde{\zeta}}(Z \leq z - \delta) + o_P(1) \leq \mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z).$$

Putting together the above inequalities, and noticing that by assumption  $\mathbb{P}(\hat{Z} \leq z) \rightarrow L_Q(z)$  for all  $z$ ,

$$L_Q(z - \delta) + o_P(1) \leq \mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) \leq L_Q(z + \delta) + o_P(1),$$

for all fixed  $z$  and  $\delta$ . By the continuity of  $L_Q(\cdot)$ , for any  $s > 0$  it is possible to choose  $\delta > 0$  small enough such that for any arbitrary  $s > 0$ ,  $|L_Q(z - \delta) - L_Q(z)| < s$  and  $|L_Q(z + \delta) - L_Q(z)| < s$ . So, for such a choice of  $\delta$ ,

$$-s + o_P(1) \leq \mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) - L_Q(z) \leq s + o_P(1).$$

By the arbitrariness of  $s$ , the above display implies that  $\mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) \rightarrow^P L_Q(z)$  for all fixed  $z$ . Since  $L_Q(\cdot)$  is continuous cdf, such a pointwise convergence can be turned into a uniform convergence, so  $\sup_{z \in \mathbb{R}} |\mathbb{P}_n^{\tilde{\zeta}}(\hat{Z} \leq z) - L_Q(z)| = o_P(1)$  (see page [van der Vaart \(1998\)](#), page 339).  $\square$

Similarly as done for Proposition 1, we now prove a more general version of Proposition 2. See Remark [A.1](#)

## Proof of Proposition 2.

(i) From Lemma [B.9](#), it is sufficient to derive a Bahadur representation for  $\sqrt{n} \cdot \mathbb{P}_n \partial_{tt} W(\beta_0) f_{W(\beta_0)}(t) \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - \hat{W}(\hat{\beta}))$ . To this end, notice that by Lemma [B.10](#) and Lemma [D.1](#),

$$\begin{aligned} &\sqrt{n} \cdot \mathbb{P}_n \partial_{tt} W(\beta_0) f_{W(\beta_0)}(t) \cdot t_n \cdot (\tilde{W}^*(\hat{\beta}^*) - \hat{W}(\hat{\beta})) = \\ &= -\sqrt{n} \cdot \mathbb{P}_n \beta_{0,2} \cdot \alpha_1(t) \cdot (\tilde{m}^* - \hat{m}_2) \cdot t_n + \sqrt{n} \cdot \mathbb{P}_n \alpha_2(t) \cdot (\tilde{m}_0^* - \hat{m}_0) \cdot t_n - \\ &\quad \mathbb{E}\{(\alpha_1(t) \cdot X_1, \alpha_1(t) \cdot m_2(X_2))\} \cdot \sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) + o_P(1) \end{aligned}$$

uniformly in  $t \in \mathbb{T}$ . Conclude by replacing in the above expression the influence-function representation of Lemma [B.10](#) and the expansion of Lemma [B.6](#).

(ii) Define the collection  $\mathbf{G} := \{g_t : t \in \mathbb{T}\}$  and the variable  $\tilde{Z}$  similarly to the proof of Proposition 1. Denote

with  $G_n^{(\xi)}$  the operator  $g \mapsto G_n^{(\xi)} g_t = \sqrt{n} P_n \tilde{\xi} \cdot g_t(\tilde{Z})$ .  $G_n^{(\xi)}$  is called bootstrap empirical process operator. For any realization of  $\{\tilde{Z}_i\}_{i=1}^n$  and  $\{\tilde{\xi}_i\}_{i=1}^n$ , the operator  $G_n^{(\xi)}$  belongs to  $A(\mathfrak{G})$ . Denote with  $G$  the Gaussian process defined in the proof of Proposition 1. Finally, denote with  $\Delta$  the operator  $f \mapsto \int |f \circ g_t|^2 \mu(dt)$  defined over the space  $A^\infty(\mathfrak{G})$ .  $\Delta$  is well defined over the space  $A^\infty(\mathfrak{G})$ . By these definitions,  $\Delta \circ G_n^{(\xi)} = \int |G_n^{(\xi)} g_t|^2 \mu(dt)$ . From Theorem 10.6 in [Kosorok \(2007\)](#)<sup>18</sup> and the proof of Proposition 2,  $G_n^{(\xi)} \rightarrow G$  over the space  $l^\infty(\mathfrak{G})$ ,  $P$ -almost surely (i.e. for  $P$ -almost every trajectory  $\{\tilde{Z}_i\}_{i=1}^n$ ). So, by the Continuous Mapping Theorem (see Theorem 18.11 in [Vaart \(1998\)](#)),  $\Delta \circ G_n^{(\xi)} \rightarrow \Delta \circ G$   $P$ -almost surely. Accordingly, since the process  $\Delta \circ G_n^{(\xi)}$  takes value in  $\mathbb{R}$ ,

$$P_n^\xi(\Delta \circ G_n^{(\xi)} \leq z) \rightarrow L(\Delta \circ G \leq z) \text{ for all } z, P\text{-almost surely,} \quad (\text{B.32})$$

where  $L$  is the cdf of the random variable  $\Delta \circ G$ , i.e.  $L$  is the cdf generated by the measure  $\Delta \circ G$ .

Now, by Lemma [B.11](#),  $G_n^{(\xi)} \rightarrow G$  over  $A(\mathfrak{G})$  with respect to the joint probability space  $\mathbb{P} \otimes P_n^\xi$ . By arguing similarly to the proof of Proposition 1, such a weak convergence implies that  $\|G_n^{(\xi)}\|_{G, \infty} = o_P(1)$ . Define now

$$R_n^*(t) := \sqrt{n} P_n \tilde{\varepsilon}^{*f} \hat{W}^{*}(\hat{\beta}^{*}) \cdot \hat{\varphi}_t \cdot \hat{t}_n - G_n^{(\xi)} g_t.$$

From point (i),  $\|R_n^*\|_{T, \infty} = o_P(1)$  so,

$$\int |S_n^* - \Delta \circ G_n^{(\xi)}| \leq \int |R_n^*(t)| \cdot |R_n^*(t) + 2G_n^{(\xi)} g_t| \mu(dt) = o_P(1).$$

So, since the above display implies a convergence over the joint space,

$$P_n^\xi(|S_n^* - \Delta \circ G_n^{(\xi)}| > \delta) = o_P(1) \text{ for all } \delta > 0. \quad (\text{B.33})$$

Since  $L(\Delta \circ G \leq \cdot)$  is a continuous function, from Eq. [\(B.33\)](#), [\(B.32\)](#) and Lemma [\(B.12\)](#),

$$\sup_{z \in \mathbb{R}} |P_n^\xi(S_n^* \leq z) - L(\Delta \circ G \leq z)| = o_P(1).$$

Since  $\hat{c}_{1-\alpha} := \inf\{c : P_n^\xi(S_n^* \leq c) \geq 1 - \alpha\}$ , it follows that (see e.g. [Li & Racine \(2006\)](#), page 213, eq. 6.40)

$$\hat{c}_{1-\alpha} \xrightarrow{P} c_{1-\alpha}, \quad (\text{B.34})$$

where  $c_{1-\alpha}$  is defined as the  $(1 - \alpha)$ -quantile of the distribution  $L(\Delta \circ G \leq \cdot)$ . Now, from Proposition 1 and the definition of the functional  $\Delta$ ,  $S_n = \Delta \circ G$ . So, because  $S_n$  takes values in  $\mathbb{R}$ ,  $P(S_n \leq z) \rightarrow L(\Delta \circ G \leq z)$  for all  $z$ . By the continuity of  $L(\Delta \circ G \leq \cdot)$ ,

$$\sup_{z \in \mathbb{R}} |P(S_n \leq z) - L(\Delta \circ G \leq z)| = o(1) \quad (\text{B.35})$$

<sup>18</sup>The same result can also be obtained from theorem 3.6.1 and Theorem 3.6.3 in [van der Vaart & Wellner \(1996\)](#), but the formulation in Theorem 10.6 in [Kosorok \(2007\)](#) is more direct for our purposes.

By Eq. (B.34) and (B.35),

$$\begin{aligned} & |P(S_n \leq \hat{c}_{1-\alpha}) - (1 - \alpha)| \leq \\ & \sup_{z \in \mathbb{R}} |P(S_n \leq z) - L(\Delta \circ G \leq z)| + |L(\Delta \circ G \leq \hat{c}_{1-\alpha}) - (1 - \alpha)| = \\ & o(1) + |L(\Delta \circ G \leq \hat{c}_{1-\alpha}) - L(\Delta \circ G \leq c_{1-\alpha})| = o_P(1), \end{aligned}$$

where I have used  $L(\Delta \circ G \leq c_{1-\alpha}) = (1 - \alpha)$ .

(iii) Define  $\mathcal{A}_c^\circ(\mathbb{T})$  as the space of the functions mapping  $\mathbb{T}$  onto  $\mathbb{R}$  that are continuous. By the proof of Proposition 1, the mapping  $g \mapsto \int |g(t)|^2 \mu(dt)$  is continuous over the set  $\mathcal{A}_c^\circ(\mathbb{T})$ . So, from Lemma C.2 and Lemma C.4, by a Continuous Mapping Theorem (see Theorem 18.11 Vaart (1998)),

$$\frac{S_n}{n} \xrightarrow{P} \int |P(Y - t t_{V(\beta^*)}(V(\beta^*))) \cdot f_{V(\beta^*)}(V(\beta^*)) \cdot \varphi| \mu(dt) \text{ and } \frac{S_n^*}{n} \xrightarrow{P} 0$$

Notice that under  $H_1$ ,  $\tilde{c} := \int |P(Y - t t_{V(\beta^*)}(V(\beta^*))) \cdot f_{V(\beta^*)}(V(\beta^*)) \cdot \varphi| \mu(dt) > 0$ . From  $S_n^* \xrightarrow{P} 0$ , it also holds that

$$P_n^\circ(S_n^* \leq n \cdot b) = o(1), \text{ for any } b > 0.$$

From the above display,  $P(S_n \leq \hat{c}_{1-\alpha}) = P(\{S_n \leq \hat{c}_{1-\alpha}\} \cap \{P_n^\circ(S_n^* \leq n \cdot b) \leq \delta\}) + o(1) = P(\{S_n \leq \hat{c}_{1-\alpha}\} \cap \{P_n^\circ(S_n^* \leq n \cdot b) \geq 1 - \delta\}) + o(1)$  for any fixed  $\delta > 0$  and  $b > 0$ . Set now  $\delta = \alpha/2$  and  $b = \tilde{c}/3$ . Since  $P(S_n/n < \tilde{c}/2) = o(1)$ , it also holds that

$$P(S_n \leq \hat{c}_{1-\alpha}) = P(\{S_n \leq \hat{c}_{1-\alpha}\} \cap \{P_n^\circ(S_n^* \leq n \cdot \tilde{c}/3) \geq 1 - \alpha/2\} \cap \{S_n/n \geq \tilde{c}/2\}) + o(1)$$

Whenever the event  $\{P_n^\circ(S_n^* \leq n \cdot \tilde{c}/3) \geq 1 - \alpha/2\}$  holds, since  $\hat{c}_{1-\alpha} = \inf\{c : P_n^\circ(S_n^* \leq c) \geq 1 - \alpha\}$ , it must be that  $n \cdot \tilde{c}/3 \geq \hat{c}_{1-\alpha}$ . Therefore, from this implication and the above display,

$$P(S_n \leq \hat{c}_{1-\alpha}) \leq P(\{S_n \leq \hat{c}_{1-\alpha}\} \cap \{n \cdot \tilde{c}/3 \geq \hat{c}_{1-\alpha}\} \cap \{S_n/n \geq \tilde{c}/2\}) + o(1) = o(1),$$

where the last inequality follows from noticing that  $\{S_n \leq \hat{c}_{1-\alpha}\} \cap \{n \cdot \tilde{c}/3 \geq \hat{c}_{1-\alpha}\} \cap \{S_n/n \geq \tilde{c}/2\} \subset \{S_n < S_n\} = \emptyset$

[Q.E.D.]

## C Power Analysis

In this Appendix, I analyze the behavior of the test under the alternative  $H_1$ .

**Lemma C.1.** *Under Assumptions A1-A6, if  $H_1$  holds,  $\hat{\beta} \xrightarrow{P} \beta^*$ .*

*Proof.* Define

$$\hat{Q}_n(\beta) = P_n(Y - \hat{t}_{\tilde{W}(\beta)}(\tilde{W}(\beta)))^2 \cdot \hat{t}_n = P_n(Y - tt_{W(\beta)}(W(\beta)))^2 + o_P(1),$$

By Lemma D.2 and Lemma A.2,  $\hat{Q}_n(\beta) = P_n(Y - tt_{W(\beta)}(W(\beta)))^2 + o_P(1)$  uniformly in  $\beta \in \mathbf{B}$ . By Assumption A1,  $tt_{W(\beta)}(W(\beta))$  satisfies a Glivenko-Cantelli property (see Theorem 19.4 in Vaart (1998)), so  $P_n(Y - tt_{W(\beta)}(W(\beta)))^2 = P(Y - tt_{W(\beta)}(W(\beta)))^2 + o_P(1)$  uniformly over  $\mathbf{B}$ . Conclude by the definition of  $\beta^*$  and Theorem 5.7 in Vaart (1998).  $\square$

**Lemma C.2.** Under Assumptions A1-A6, under  $H_1$ , uniformly over  $\mathbf{T}$

$$P_n \tilde{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{q} \cdot \hat{t}_n = P(Y - tt_{V(\beta^*)}(V(\beta^*))) \cdot f_{V(\beta^*)}(V(\beta^*)) \cdot \varphi(X\tilde{t}) + o_P(1).$$

*Proof.* By a reasoning similar to the previous lemma, together with  $\hat{\beta} = \beta^* + o_P(1)$ ,

$$\begin{aligned} & P_n(Y - \hat{t}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta}))) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(W(\hat{\beta})) \cdot \hat{q} \cdot \hat{t}_n = \\ & P(Y - tt_{W(\beta^*)}(W(\beta^*))) \cdot f_{W(\beta^*)}(W(\beta^*)) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ . For the bias correction term, notice that by the uniform-in- $\beta$  convergence results of Lemma D.2,

$$\begin{aligned} & P_n \hat{B}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{t}_n = \\ & P_n(tt_{\hat{\beta}|W(\hat{\beta})}(W(\hat{\beta})) - tt_{W(\beta^*)}(W(\beta^*))) \cdot t_n \cdot f_{W(\beta^*)}(W(\beta^*)) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ . Since by definition  $tt_{\beta_1|W(\beta_2)}(w) = E\{tt_{W(\beta_1)}(W(\beta_1)) | W(\beta_2) = w\}$ , it also holds that  $\hat{t}_{\hat{\beta}|W(\hat{\beta})} = tt_{W(\hat{\beta})}$ . Accordingly,  $P_n \hat{B}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta})}(\tilde{W}(\hat{\beta})) \cdot \hat{t}_n = o_P(1)$  uniformly over  $\mathbf{T}$ .  $\square$

**Lemma C.3.** Under Assumptions A1-A6, under  $H_1$ ,  $\hat{\beta} \xrightarrow{P} \beta^*$ .

*Proof.* From Lemma D.2, Lemma C.1, and Lemma A.2,

$$\begin{aligned} & P_n(\hat{t}_{\hat{W}(\beta_1)}(\hat{W}(\beta_1)) + \xi \cdot (Y - \hat{t}_{\hat{W}(\beta_1)}(\hat{W}(\beta_1))) - \hat{t}_{\tilde{W}^*(\beta_2)}^*(\tilde{W}(\beta_2)))^2 \cdot \hat{t}_n = \\ & P_n(tt_{W(\beta_1)}(W(\beta_1)) + \xi \cdot (Y - tt_{W(\beta_1)}(W(\beta_1))) - tt_{\beta^*|W(\beta_2)}(W(\beta_2)))_2 + o_P(1), \end{aligned}$$

uniformly in  $(\beta_1, \beta_2) \in \mathbf{B} \times \mathbf{B}$ . By Assumption A2,  $tt_{W(\beta)}(W(\beta))$  satisfies a Glivenko Cantelli property (see Theorem 19.4 in van der Vaart (1998)), so

$$\begin{aligned} & P_n(tt_{W(\beta_1)}(W(\beta_1)) + \xi \cdot (Y - tt_{W(\beta_1)}(W(\beta_1))) - tt_{\beta^*|W(\beta_2)}(W(\beta_2)))_2 = \\ & P(tt_{W(\beta_1)}(W(\beta_1)) + \xi \cdot (Y - tt_{W(\beta_1)}(W(\beta_1))) - tt_{\beta^*|W(\beta_2)}(W(\beta_2)))_2 + o_P(1) \end{aligned}$$

uniformly in  $(\beta_1, \beta_2) \in \mathbf{B} \times \mathbf{B}$ . By Assumption A1 and Lemma C.1,

$$\begin{aligned} & P(tt_{W(\hat{\beta})}(W(\hat{\beta})) + \zeta \cdot (Y - tt_{W(\hat{\beta})}(W(\hat{\beta}))) - tt_{\beta^*|W(\beta)}(W(\beta_2)))^2 = \\ & P(tt_{W(\beta^*)}(W(\beta^*)) + \zeta \cdot (Y - tt_{W(\beta^*)}(W(\beta^*))) - tt_{\beta^*|W(\beta_2)}(W(\beta_2)))^2 + o_P(1) \\ & =: Q^*(\beta_2) + o_P(1) \end{aligned}$$

uniformly in  $\beta_2 \in \mathbf{B}$ . By the previous three displays,  $\hat{Q}_n^*(\beta) = Q^*(\beta) + o_P(1)$  uniformly over  $\mathbf{B}$ .

Notice now that, since  $E\zeta^2 = 1$  and  $E\zeta = 0$ ,  $Q^*(\beta) = P(tt_{W(\beta^*)}(W(\beta^*)) - tt_{\beta^*|W(\beta)}(W(\beta)))^2 + P(Y - tt_{W(\beta^*)}(W(\beta^*)))^2$ . Hence,

$$\arg \min_{\beta \in \mathbf{B}} Q^*(\beta) = \{\beta \in \mathbf{B} : tt_{W(\beta^*)}(W(\beta^*)) = tt_{\beta^*|W(\beta)}(W(\beta))\} \text{ } P\text{-almost surely } \}.$$

Differentiating both sides of  $tt_{W(\beta^*)}(W(\beta^*)) = tt_{\beta^*|W(\beta)}(W(\beta))$  wrt to  $(x_1, m_2(x_2))$ , yields  $\beta = \beta^*$ . Hence,  $\arg \min_{\beta \in \mathbf{B}} Q^*(\beta) = \{\beta^*\}$ . Conclude by recalling that  $\hat{\beta}^* := \arg \min_{\beta \in \mathbf{B}} \hat{Q}_n^*(\beta)$  and using Theorem 5.7 in [van der Vaart \(1998\)](#).  $\square$

**Lemma C.4.** Under Assumption A1-A6, if  $H$  holds,  $P_n \tilde{\varepsilon}_n^* \cdot f_{\tilde{W}(\hat{\beta}^*)}^* (\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \hat{=} o(1)$  uniformly over  $\mathbf{T}$ .

*Proof.* From Lemma D.2 and Lemma A.2,

$$\begin{aligned} & P_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}^* (\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \hat{=} \\ & P_n(tt_{\hat{\beta}^*|W(\hat{\beta})}(W(\hat{\beta})) - tt_{W(\hat{\beta})}(W(\hat{\beta}))) \cdot f_{W(\hat{\beta})}(W(\hat{\beta})) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ . Since  $tt_{\hat{\beta}^*|W(\hat{\beta})} = tt_{W(\hat{\beta})}$  (see the proof of Lemma C.2),

$$P_n \hat{B}_{\hat{W}(\hat{\beta})}(\hat{W}(\hat{\beta})) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}^* (\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \hat{=} o(1)$$

uniformly over  $\mathbf{T}$ . Now, by Lemma D.2 and Lemma A.2,

$$\begin{aligned} & P_n(Y - tt_{W(\hat{\beta})}(W(\hat{\beta})) + \zeta \cdot (Y - tt_{W(\hat{\beta})}(W(\hat{\beta}))) - tt_{\hat{\beta}^*|W(\hat{\beta}^*)}(W(\hat{\beta}^*))) \cdot \hat{f}_{\tilde{W}(\hat{\beta}^*)}^* (\tilde{W}(\hat{\beta}^*)) \cdot \varphi_t \hat{=} \\ & P_n(tt_{W(\hat{\beta})}(W(\hat{\beta})) + \zeta \cdot (Y - tt_{W(\hat{\beta})}(W(\hat{\beta}))) - tt_{\hat{\beta}^*|W(\hat{\beta}^*)}(W(\hat{\beta}^*))) \cdot f_{W(\hat{\beta}^*)}(W(\hat{\beta}^*)) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ . By Assumption A2, a mean-value expansion, Lemma C.3, and Lemma C.1 yield

$$\begin{aligned} & P_n(tt_{V(\hat{\beta})}(V(\hat{\beta})) + \zeta \cdot (Y - tt_{V(\hat{\beta})}(V(\hat{\beta}))) - tt_{\hat{\beta}^*|V(\hat{\beta}^*)}(V(\hat{\beta}^*))) \cdot f_{V(\hat{\beta}^*)}(V(\hat{\beta}^*)) \cdot \varphi_t = \\ & P_n(\zeta \cdot (Y - tt_{V(\hat{\beta}^*)}(V(\hat{\beta}^*)))) \cdot f_{V(\hat{\beta}^*)}(V(\hat{\beta}^*)) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbf{T}$ , where I have used  $tt_{\hat{\beta}^*|V(\hat{\beta}^*)} = tt_{V(\hat{\beta}^*)}$ . The compactness of  $\mathbf{T}$  and the boundedness of  $\tilde{X}$  ensure



a Glivenko-Cantelli property, thus

$$\begin{aligned} & P_n(\zeta \cdot (Y - tt_{V(\beta^*)}(V(\beta^*))) \cdot f_{V(\beta^*)}(V(\beta^*)) \cdot \varphi_t = \\ & P(\zeta \cdot (Y - tt_{V(\beta^*)}(V(\beta^*))) \cdot f_{V(\beta^*)}(V(\beta^*)) \cdot \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $T$ . Conclude by recalling that  $E\zeta = 0$ . □

## D Auxiliary Results

The following lemmas are proved in the [Supplementary Material](#) of this paper.

**Lemma D.1.** For  $j = 0, 2$ , define  $d_{n,j} := \frac{\log n}{n h_j^{p_j}} + h_j^r$ . Under Assumptions A1-A5,

- (i)  $\|(\hat{m}_j - m_j) \cdot t_n\|_\infty = O_P(\frac{d_{n,j}}{\tau_n})$  and  $\|(\hat{m}_j^* - m_j) \cdot t_n\|_\infty = O_P(\frac{d_{n,j}}{\tau_n^2})$ ;
- (ii)  $\|(\hat{T}_{\hat{m}_j}/\hat{f}_j - m_j) \cdot t_n\|_\infty = O_P(\frac{d_{n,j}}{\tau_n^2})$ ;
- (iii)  $\frac{\hat{T}_j - \hat{T}_{\hat{m}_j}}{\hat{f}_j} \cdot t_n - \frac{\hat{T}_j - \hat{T}_{\hat{m}_j}}{\hat{f}_j} \cdot \hat{t}_\infty = o_P(n^{-1/2})$  and  $\frac{\hat{T}_j^* - \hat{T}_j}{\hat{f}_j} \cdot t_n - \frac{\hat{T}_j^* - \hat{T}_j}{\hat{f}_j} \cdot \hat{t}_\infty = o_P(n^{-1/2})$

For  $C_{0,n} := C(Z_n)$  and  $C_{2,n} := C(X_{2,n})$ ,

- (iv)  $P(\hat{m}_j \in C_{j,n} \xrightarrow{\Sigma} 1, P(\frac{\hat{T}_{\hat{m}_j}}{\hat{f}_j} \in C_{j,n} \xrightarrow{\Sigma} 1, P(\hat{m}_j^* \in C_{j,n} \xrightarrow{\Sigma} 1.$

**Lemma D.2.** Under Assumption A1-A6,

- (i)  $\sup_{\beta \in B} \sup_{x \in \tilde{X}} |\hat{f}_{\tilde{W}(\beta)}(\tilde{w}(\beta)) - tt_{W(\beta)}(w(\beta))| \cdot t_n(x \sim) = o_P(n^{-1/4})$ ;
- The same rate also holds for  $\hat{f}_{\tilde{W}(\beta)}(\tilde{w}(\beta))$ ,  $\hat{f}_{\tilde{W}(\beta)}(\tilde{w}(\beta))$ , and  $\hat{f}_{\tilde{W}(\beta)}(\tilde{w}(\beta))$ .

- (ii)  $\sup_{\beta \in B} \sup_{x \in \tilde{X}} |\hat{f}_{\tilde{W}(\beta)}(\tilde{w}^*(\beta)) - f_{W(\beta)}(w(\beta))| \cdot t_n(x \sim) = o_P(n^{-1/4})$ ;

Define  $tt_{\beta_1|W(\beta_2)}(w) := E\{tt_{W(\beta_1)}(W(\beta_1)) \mid W(\beta_2) = w\}$ ; then,

- (iii)  $\sup_{\beta \in B} \sup_{x \in \tilde{X}} |\hat{f}_{\tilde{W}^*(\beta)}(\tilde{w}^*(\beta)) - tt_{\beta|W(\beta)}(w(\beta))| \cdot t_n(x \sim) = o_P(1)$

- (iv)  $\sup_{\beta \in B} \sup_{x \in \tilde{X}} |\frac{\hat{T}_{\hat{G}_{\tilde{W}(\beta)}}(\hat{w}(\beta))}{\hat{f}_{\tilde{W}(\beta)}(\hat{w}(\beta))} - tt_{\beta|W(\beta)}(w(\beta))| \cdot t_n(\tilde{x}) = o_P(n^{-1/4})$ ;

The above results also hold by replacing  $t_n$  with  $\hat{t}_n$ .

**Lemma D.3.** Under Assumptions A1-A6, if  $H_0$  holds and  $\sqrt{n}(\hat{\beta} - \beta_0) = O_P(1)$

- (i)  $\sup_w |T_{\tilde{W}(\hat{\beta})}(\hat{w}) - T_{W(\beta_0)}(w)| \cdot t_{n,W(\beta_0)}^{(\eta/3)}(w) = o_P(n^{-1/4} \cdot \tau_n^{2s+4})$ ,
- (ii)  $\sup_w |\partial T_{\tilde{W}(\hat{\beta})}(\hat{w}) - \partial T_{W(\beta_0)}(w)| \cdot t_{n,W(\beta_0)}^{(\eta/3)}(w) = o_P(n^{-1/4} \cdot \tau_n^{2s+4})$ .

The same rate also holds for  $\hat{f}_{\tilde{W}(\hat{\beta})}(\hat{w})$ ,  $\hat{f}_{\tilde{W}(\hat{\beta})}(\hat{w})$ ,  $\hat{f}_{\tilde{W}(\hat{\beta})}(\hat{w})$ ,  $\hat{f}_{\tilde{W}(\hat{\beta})}(\hat{w})$ ,  $\hat{f}_{\tilde{W}(\hat{\beta})}(\hat{w})$ , and  $\hat{T}_{\hat{w}(\hat{\beta})}$ .

If moreover  $\sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) = O_P(1)$ ,

$$(iii) \sup_w |T_{\tilde{W}^*(\hat{\beta}^*)}^*(w) - T_{W(\hat{\beta})}(w)| \cdot t_{n, W(\hat{\beta})}^{(\eta/3)}(w) = o_P(n^{-1/4} \cdot \tau_n^{2s+4}),$$

$$(iv) \sup_w |\partial T_{\tilde{W}^*(\hat{\beta}^*)}^*(w) - \partial T_{W(\hat{\beta})}(w)| \cdot \tau_n^{2s+4} = o_P(n^{-1/4} \cdot \tau_n);$$

The same result holds for  $f_{\tilde{W}^*(\hat{\beta}^*)}$ .

**Lemma D.4.** Under Assumption A1-A6, if  $H_0$  holds and  $\sqrt{n} \cdot (\hat{\beta} - \beta_0) = O_P(1)$ ,

$$(i) P(f_{\tilde{W}(\hat{\beta})}, T_{\tilde{W}(\hat{\beta})}, f_{\hat{W}(\hat{\beta})}, T_{\hat{W}(\hat{\beta})}, T_{it_{\tilde{V}(\hat{\beta})}}, T_{it_{\hat{V}(\hat{\beta})}} \in C(W_n)) \rightarrow 1.$$

If moreover  $\sqrt{n} \cdot (\hat{\beta}^* - \hat{\beta}) = O_P(1)$ ,

$$(ii) P(f_{\tilde{W}^*(\hat{\beta}^*)}, T_{\tilde{W}^*(\hat{\beta}^*)}^* \in C(W_n)) \rightarrow 1.$$

# Chapter 2: A Nonparametric Encompassing Test

Elia Lapenta and Pascal Lavergne\*

## Abstract

According to the encompassing principle, a model  $M_1$  encompasses a model  $M_2$  if  $M_1$  can explain the results of  $M_2$ . The encompassing tests in the current literature either rely on parametric functional forms or, when relying on nonparametric specifications, they condition the analysis on fixed values of the explanatory variables. In this paper, we provide a nonparametric encompassing test. Our procedure does not rely on neither functional forms nor on specific values of the explanatory variables. We propose a statistic computed according to an  $L_2$ -boosting algorithm. This procedure allows to obtain a good robustness with respect to the choice of the tuning parameter. We propose to simulate the critical values by a wild-bootstrap procedure and prove its validity. In a Monte-Carlo simulation study, we show the attractive features of our test.

**Keywords:** Encompassing, Model Selection, Nonparametric Model, Bootstrap,  $L_2$  Boosting.

**JEL Classification:** C01, C12, C14

## 1 Introduction

Encompassing is a well established principle in econometrics. It allows the researcher to choose between two competing models explaining the behavior of the same response variable. Essentially, a model  $M_1$  encompasses a model  $M_2$  if  $M_1$  can explain the results of  $M_2$ . This is a natural principle for choosing between two competing theories: a new theory can replace an older one not only if it explains a new phenomenon, but also if it can explain the results of the older theory.

The encompassing principle has been introduced in econometrics by the work of [Mizon & Richard \(1986\)](#), [Hendry & Richard \(1987\)](#), [Gourieroux \*et al.\* \(1983\)](#), [Florens \*et al.\* \(1996\)](#). [Gourieroux & Monfort \(1995\)](#) have developed encompassing tests for parametric models, while [Florens \*et al.\* \(1996\)](#) have extended the concept of encompassing to a Bayesian framework. An extensive

---

\*Toulouse School of Economics.

survey of the encompassing principle is provided in [Bontemps & Mizon \(2008\)](#). [Bontemps et al. \(2008\)](#) propose different encompassing tests, concerning both parametric and nonparametric setups. Tests in the current literature either rely on parametric functional forms or, when relying on nonparametric specifications, they condition the analysis on fixed values of the explanatory covariates. Accordingly, the results obtained can be considered as conditional on either the specific parametric classes or on the values of the explanatory variables.

This paper contributes to the literature on the encompassing principle by providing a test that is fully nonparametric and is not conditional on fixed values of explanatory covariates. The test statistic we propose is based on a continuum of moments and is built according to an  $L_2$  boosting algorithm. Such an algorithm has been originally proposed in the machine learning literature to debias the estimators of nonparametric objects. Since our test is fully nonparametric, we use the  $L_2$  boosting procedure to recursively correct the nonparametric estimators employed. This algorithm turns out to be effective for two main reasons. First, it considerably enlarges the spectrum of bandwidths and kernels admissible for the test. Second, it makes the procedure more robust to the bandwidth choice. We show that, under the null, the statistic proposed converges to a functional of a Gaussian process which depends on unknown features of the data. So, to obtain the critical value we propose a wild-bootstrap procedure and prove its validity in the presence of boosting iterations. In a Monte Carlo simulation study we explore the merits of our procedure and show the robustness with respect to the choice of the smoothing parameter.

The reminder of the paper goes as follows. Section [2](#) formalizes the encompassing idea and draws some links between the nonparametric context considered in this paper and the one analyzed in other works. It moreover provides some interesting features of the encompassing principle. Section [3](#) constructs the test, Section [4](#) introduces the  $L_2$  boosting estimators, while the following Section [5](#) sets up the assumptions and obtains the asymptotic behavior of the statistic. Since the asymptotic distribution under the null depends on unknown features of the DGP, in Section [6](#) we propose a Wild-Bootstrap procedure and prove its validity. Section [7](#) provides evidence about the small-sample behavior of our test. Finally, Section [8](#) concludes. Appendix [A](#), [B](#), and [C](#) contain all the technical proofs.

## 2 The Encompassing Principle

Let  $\{Y_i, W_i, X_i\}_{i=1}^n$  be an iid sample from a population, and imagine to have two competing models seeking to explain the behavior of  $Y$ . For example, model  $M_1$  might explain the behavior of  $Y$  by the covariates  $W$  and model  $M_2$  by the regressors  $X$ . According to the encompassing principle, one theory encompasses the other if the former can explain the results of the latter. In the present context, a model seeking to explain  $Y$  with a specific vector of regressors can be formalized by the

function of these regressors that is the best approximation of  $Y$ . We can consider the  $L_2$  distance to measure the quality of such approximation. Since  $M_1$  seeks to explain  $Y$  by the regressors  $W$ , it can be defined as  $M_1 := L^2(W)$ , where  $L^2(W)$  is the space of the square integrable functions of  $W$ . Similarly, define  $M_2 := L^2(X)$ , with  $L^2(X)$  denoting the space of square integrable functions of  $X$ . The best approximations of  $Y$  resulting from model  $M_1$  and  $M_2$ , respectively, are obtained as

$$r_1 := \arg \min_{g \in M_1} \|Y - g(W)\|_2 \text{ and } r_2 := \arg \min_{g \in M_2} \|Y - g(X)\|_2, \quad (1)$$

where  $\|\cdot\|$  denotes the  $L^2$ - norm of square-integrable random variables, i.e.  $\|Z\|^2 := \int z^2 P(dz)$  for any random element  $Z$ . By the previous definitions,  $r_1$  is the  $L^2$  projection of  $Y$  onto  $L^2(W)$ , and  $r_2$  is the  $L^2$  projection of  $Y$  onto  $L^2(X)$ . Hence,

$$r_1(W) = E\{Y | W\} \text{ and } r_2(X) = E\{Y | X\}.$$

According to the encompassing principle, model  $M_1$  encompasses model  $M_2$  if  $M_1$  can explain the results of  $M_2$ . Since  $r_1$  is the “explanation” of  $Y$  resulting from model  $M_1$ , and  $r_2$  is the “explanation” of  $Y$  resulting from model  $M_2$ , we can say that model  $M_1$  encompasses model  $M_2$  if  $r_2$  can be “obtained” from  $r_1$ . Formally,

$$r_2 = \arg \min_{g \in M_2} \|r_1(W) - g(X)\|_2.$$

Hence, using again the definition of the  $L^2$  projection,  $M_1$  encompasses  $M_2$  if and only if

$$E\{r_1(W) | X\} = r_2(X). \quad (2)$$

## 2.1 Relations with the definition of Encompassing in **Gourieroux and Monfort (1995)**

The above formulation of the encompassing principle can be considered as a nonparametric generalization of the definition provided in [Gourieroux & Monfort \(1995\)](#) (GM, now henceforth). In particular, the authors start from two competing models aiming to explain the distribution of a variable  $Y$ . Then, they define a pseudo-true value for a model as the element of the model that is the closest to the true distribution of  $Y$  in terms of Kullback-Leibler distance. In particular, denote with  $f_0$  the true distribution of  $Y$ . Each model in GM is respectively defined by the following collections

$$\tilde{M}_1 := \{g_1(\cdot, \alpha_1) : \alpha_1 \in A_1\} \text{ and } \tilde{M}_2 := \{g_2(\cdot, \alpha_2) : \alpha_2 \in A_2\},$$

where  $g_j(\cdot, \alpha_j)$  for  $j = 1, 2$  denotes a specific distribution indexed by the parameter  $\alpha_j$ , and  $A_1, A_2$  are two generic sets. Notice that the parameters  $\alpha_1, \alpha_2$  are not restricted to be real vectors. The pseudo true value of  $\alpha_j$  is therefore defined as

$$\alpha_j^* = \arg \min_{\alpha_j \in A_j} d_{KL}(f_o, g_j(\cdot, \alpha_j)),$$

where  $d_{KL}(\cdot, \cdot)$  denotes the Kullback-Liebler distance between two densities. GM also define the element of model  $\tilde{M}_2$  that is the closest possible to an element of model  $\tilde{M}_1$ :

$$b(\alpha_1) := \arg \min_{\alpha_2 \in A_2} d_{KL}(g_1(\cdot, \alpha_1), g_2(\cdot, \alpha_2)),$$

for  $\alpha_1 \in A_1$ . According to the definition of encompassing in GM, model  $\tilde{M}_1$  encompasses model  $\tilde{M}_2$  if and only if

$$\alpha_2^* = b(\alpha_1^*).$$

In other words, from the above definition  $\tilde{M}_1$  encompasses  $\tilde{M}_2$  if the pseudo-true value of the latter can be “obtained” from the pseudo-true value of the former.

The definition considered in this paper is very similar to the one in GM. We are just replacing the KL distance with the  $L^2$  norm, and considering nonparametric regression functions instead of densities. From the definitions in Eq. 1,  $r_1$  and  $r_2$  can be viewed as the pseudo-true values obtained from the models  $M_1$  and  $M_2$ , respectively. Also, the element of  $M_2$  that is the closest to a generic  $g_1 \in M_1$  is defined as

$$\tilde{b}(g_1) = \arg \min_{g \in M_2} ||g_1(W) - g_2(X)||_2.$$

In the definition of GM, Model  $\tilde{M}_1$  encompasses model  $\tilde{M}_2$  if and only if  $\tilde{b}(r_1) = r_2$ . This exactly coincides with the definition we have provided in Eq. 2.

## 2.2 Some features of the Encompassing Principle

The formulation provided in Eq. 2 allows us to obtain some intuitive features of the encompassing principle.

First, if model  $M_1$  encompasses model  $M_2$ , the explanation of  $Y$  provided by  $M_1$  will be more “accurate” than the explanation of  $Y$  provided by  $M_2$ . To provide a formal proof, define  $\langle f, g \rangle := \int f(z) \cdot g(z) P(dz)$ , where  $z = (y, x_1, x_2)$ . So,  $\langle \cdot, \cdot \rangle$  denotes the scalar product in the  $L^2(Z)$  space.

**Proposition 2.1.** *If Model  $M_1$  encompasses model  $M_2$ , then  $\|Y - r_1(W)\| \leq \|Y - r_2(X)\|$ .*

*Proof.*  $\|Y - r_2(X)\|^2 = \|Y - r_1 + r_1 - r_2\|^2 = \|Y - r_1\|^2 + 2\langle Y - r_1, r_1 - r_2 \rangle + \|r_1 - r_2\|^2$ . Since  $r_1$  is the projection of  $Y$  onto  $L^2(W)$ ,  $\langle Y - r_1, r_1 \rangle = 0$ . By definition,  $r_2$  is the projection of  $Y$  onto  $L^2(X)$ . Also, since  $M_1$  encompasses  $M_2$ ,  $r_2$  is equal to the projection of  $r_1$  onto  $L^2(X)$ . Thus,  $\langle Y - r_1, r_2 \rangle = \langle Y, r_2 \rangle - \langle r_1, r_2 \rangle = \langle r_2, r_2 \rangle - \langle r_2, r_2 \rangle = 0$ . This implies that  $\langle Y - r_1, r_1 - r_2 \rangle = 0$ , and hence  $\|Y - r_2(X)\|^2 = \|Y - r_1\|^2 + \|r_1 - r_2\|^2 \geq \|Y - r_1\|^2$ .  $\square$

Second, if the two models mutually encompass each other, they give rise to the same “theory” or “explanation”, as the following Proposition shows.

**Proposition 2.2.** *If  $M_1$  encompasses  $M_2$ , and also  $M_2$  encompasses  $M_1$ , then  $r_1(W) = r_2(X)$   $P$ -almost surely.*

*Proof.* Since the two models are mutually encompassing,  $\|Y - r_1(W)\| \leq \|Y - r_2(X)\|$  and  $\|Y - r_2(X)\| \leq \|Y - r_1(W)\|$ , so that  $\|Y - r_1(W)\| = \|Y - r_2(X)\|$ . By proceeding as in the proof of Proposition 2.1,  $\|Y - r_2(X)\|^2 = \|Y - r_1\|^2 + \|r_1 - r_2\|^2$ , so that  $\|r_1 - r_2\| = 0$ .  $\square$

## 2.3 The Encompassing Principle and nested models

The encompassing principle considered here is linked with the principles of nested models and nonparametric significance of the regressors. Imagine that the regressors  $X$  are not significant for  $Y$  once we control for  $W$ :

$$E\{Y | W, X\} = E\{Y | W\}.$$

By applying the conditional expectation  $E\{\cdot | X\}$  to both sides of the above display and then using the Law of Iterated expectation, together with the definitions of  $r_1$  and  $r_2$ ,

$$r_2(X) = E\{r_1(W) | X\},$$

So, if the regressors  $X$  are not significant in the nonparametric regression of  $Y$  onto  $(W, X)$ , model  $M_1$  will encompass model  $M_2$ . The other direction of this implication, however, does not hold, in the sense that if model  $M_1$  encompasses model  $M_2$ , then it is not necessarily true that the covariates  $X$  are not significant in the nonparametric regression of  $Y$  onto  $(W, X)$ . This is highlighted by the following proposition.

**Proposition 2.3.** Let  $r_1(W) := E\{Y | W\}$  and  $r_2(X) := E\{Y | X\}$ . Then, model  $M_1$  encompasses  $M_2$  if and only if  $E\{Y | W, X\} = r_1(W) + g(W, X)$ , with  $E\{g(W, X) | W\} = E\{g(W, X) | X\} = 0$ .

*Proof.* Assume that  $M_1$  encompasses  $M_2$ . Notice that we can always write  $Y = r_1(W) + \varepsilon$ . So,  $E\{Y | W, X\} = r_1(W) + E\{\varepsilon | W, X\} =: r_1(W) + g(W, X)$ . Applying  $E\{\cdot | W\}$  to both sides of the previous equation, using the Law of Iterated expectation, and the definition of  $r_1(W)$ ,  $E\{g(W, X) | W\} = 0$ . Similarly, by applying  $E\{\cdot | X\}$ , and then using the Law of iterated expectation and definition of  $r_2$ ,  $r_2(X) = E\{r_1(W) | X\} + E\{g(W, X) | X\}$ . Since  $M_1$  encompasses  $M_2$ , by Eq. 2 and the previous display we must have  $E\{g(W, X) | X\} = 0$ . The other direction of the proposition is immediately proved by the Law of iterated expectations and the definitions of  $r_1$  and  $r_2$ .  $\square$

The previous result highlights that if Model  $M_1$  encompasses model  $M_2$ , we must not necessarily have  $E\{Y | W, X\} = E\{Y | W\}$ . As long as  $g(W, X)$  is not a.s. equal to zero,  $M_1$  encompasses  $M_2$  but  $X$  is a significant covariate in  $E\{Y | W, X\}$ . Thus, the use of a nonparametric significance test checking the restriction  $E\{Y | W, X\} = E\{Y | W\}$  is misleading for the encompassing assumption.

The following counterexample provides a concrete functional form for  $g$  in the last proposition. It also shows a DGP where  $X$  is significant in the expectation  $E\{Y | W, X\}$ , model  $M_1$  encompasses  $M_2$ , but the models do not mutually encompass each other.

**Counterexample.** Let  $Y = (1 + W) + g(W, X) + \eta$ , with  $\eta \perp (W, X)$  and  $g(W, X) = (W^2 - 1/12) \cdot X$ . Define

$$f_{W,X}(w, x) := 1\{w \in [-0.5, 0]\} \cdot \phi(x) + 1\{w \in [0, 0.5]\} \cdot \psi(x),$$

where  $\phi$  and  $\psi$  are two densities such that  $\int x\phi(x)dx = 0$  and  $\int x\psi(x)dx = 0$ .  $f_{W,X}$  is a density function with respect to the Lebesgue measure. Indeed,  $f_{W,X} \geq 0$  and  $\int f_{W,X}(w, x)dwd x = 1$ . Assume that  $(W, X) \sim f_{W,X}$ . The marginal densities are given by

$$f_W(w) = 1\{w \in [-0.5, 0.5]\}, \text{ i.e. the uniform distribution,}$$

$$f_X(x) = 0.5\phi(x) + 0.5\psi(x).$$

The conditional densities write as

$$f_{W|X}(w|x) = 2 \cdot \frac{1\{w \in [-0.5, 0]\} \cdot \phi(x) + 1\{w \in [0, 0.5]\} \cdot \psi(x)}{\phi(x) + \psi(x)}$$

$$f_{X|W}(x|w) = 1\{w \in [-0.5, 0]\} \cdot \phi(x) + 1\{w \in [0, 0.5]\} \cdot \psi(x), \text{ for } w \in [-0.5, 0.5].$$



From the above conditional densities, it is easy to show that  $E\{W^2|X\} = 1/12$ , and  $E\{X|W\} = 0$ . Hence,  $E\{g(W, X)|W\} = E\{g(W, X)|X\} = 0$ . By the previous proposition, this implies that  $M_1$  encompasses  $M_2$ . Clearly, since  $g(W, X)$  is not a.s. equal to zero,  $X$  is significant in the expectation  $E\{Y|W, X\}$ .

Finally,  $r_1(W) = 1 + W$ , and by using the expression of  $f_{X|W}$  obtained above,  $r_2(X) = E\{Y|X\} = E\{(1 + W)|X\} = \frac{\psi(X)}{\psi(X) + \phi(X)}$ . Now, by Proposition 2.2, if  $M_2$  encompasses  $M_1$  (i.e. the two models are mutually encompassing), then  $r_1(W) = r_2(X)$  a.s. Therefore, since  $r_1$  is one-to-one,  $W$  must be an exact function of  $X$ . But this is not satisfied by the density  $f_{W,X}$ .

### 3 Construction of a Nonparametric Encompassing Test

By the definition in Eq. 1, the null hypothesis is set as

$$H_0 : E\{r_1(W)|X\} = r_2(X),$$

where  $r_1(W) := E\{Y|W\}$ ,  $r_2(X) := E\{Y|X\}$ , with  $W \in \mathbb{R}^d$ ,  $X \in \mathbb{R}^{d_X}$ . For the ease of notation, let us denote by  $r$  the regression  $r_1$ . Using the linearity of the conditional expectation,  $H_0$  can be equivalently written as

$$H_0 : E\{Y - r(W) | X\} = 0. \quad (3)$$

We test  $H_0$  against its logical complement,  $H_1 : H_0^c$ . The above display contains a conditional moment restriction. We will transform such a conditional moment into a continuum of unconditional moments. By the results in [Bierens \(1982\)](#), [Bierens & Ploberger \(1997\)](#), and [Stinchcombe & White \(1998\)](#),

$$H_0 : E\{(Y - r(W)) \cdot \varphi_t(X)\} = 0 \text{ for all } t \in T,$$

where  $T$  compact subset of  $\mathbb{R}^{d_X}$  containing the origin, while  $\varphi_t(\cdot) = \phi(t \cdot)$ , with  $\phi$  an analytic non-polynomial function. To simplify notation, define

$$\varepsilon = Y - r(W).$$

$\hat{\varepsilon} = Y - \hat{r}(W)$  will denote an estimator of  $\varepsilon$ . Also, let  $P = \frac{1}{n} \sum_{i=1}^n \delta_{Z_i}$  with  $Z_i = (Y_i, W_i, X_i)$ , and  $\delta$  denoting the Dirac measure. For any function  $g \in L^1(\mathbb{Z})$ ,  $P g(Z) := \frac{1}{n} \sum_{i=1}^n g(Z_i)$ . The

statistic we propose is an Integrated Conditional Moment test,

$$S_n = \int ||\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{\varphi}_t||^2 \mu(dt),$$

where  $\mu$  is a measure absolutely continuous with respect to the Lebesgue measure and  $|| \cdot ||$  is the Euclidean norm on  $\mathbb{C}$ . To obtain a simple expression for our statistic, we could set  $\varphi_t(\cdot) = \exp(i t \cdot)$ , where  $i = \sqrt{-1}$ . In such a case,  $S_n$  can be expressed as a weighted double sum of the estimated residuals, as

$$S_n = \frac{1}{n} \sum_{i,j} \hat{\varepsilon}_i \hat{\varepsilon}_j \phi_\mu(X_i - X_j),$$

where  $\phi_\mu$  is the characteristic function of the measure  $\mu$ . Setting  $\varphi_t(\cdot) = \exp(i t \cdot)$  is not necessary in our proofs which hold for a general weighing function, but such specific choice of  $\varphi_t$  has the advantage of delivering a simple closed form expression for  $S_n$ .

The statistic  $S_n$  is very similar to the one proposed by [Delgado & Manteiga \(2001\)](#). There are however some differences. First, the authors use locally constant kernel estimation in the construction of their test. To simplify the proofs for the asymptotic behavior, [Delgado & Manteiga \(2001\)](#) multiply “both sides” of their null hypothesis by a density allowing them to avoid a random denominator. They therefore use a leave-one out estimator and obtain a  $U$ -process form for the empirical process at the basis of their statistic. Differently, due to the formulation of  $H_0$  in Eq. 3, we cannot multiply both sides of  $H_0$  by a density to avoid a random denominator. Second, to compute the estimated residuals  $\{\hat{\varepsilon}_i\}_{i=1}^n$  we will modify the kernel estimators by implementing an  $L_2$ -boosting procedure. This allows us to handle the bias in  $S_n$  coming from the nonparametric estimation of  $r_1$ . To better explain this point, let us consider the estimation of the residual  $\varepsilon$  by kernel methods, i.e. let us assume that the function  $r_1$  is estimated by a locally constant kernel regression. In specification testing it is generally assumed that the bandwidth  $h$  and the kernel order  $\bar{r}$  must satisfy  $nh^{2\bar{r}} = o(1)$ , see e.g. [Delgado & Manteiga \(2001\)](#) or [Escanciano et al. \(2014\)](#). This condition allows to handle the bias terms appearing in the expansion of the empirical process  $\sqrt{n} \mathbf{P}_n \hat{\varepsilon} \hat{\varphi}_t(X)$ . It however restricts the set of bandwidths and kernels admissible for the test. For example, the bandwidth coming from the minimization of the mean-square error cannot be used for the implementation of the test. Moreover, to ensure the condition  $nh^{2\bar{r}} \rightarrow 0$ , a high-order kernel must be selected when the dimension of  $X$  is larger than 1. Since high-order kernels are relatively irregular functions, this will inflate the small-sample variance of the kernel estimators, impacting negatively on the performance of the test. The boosting procedure we implement allows us to alleviate these problems.

---

**Algorithm 1**  $L_2$  Boosting

---

- Initialize with estimates  $\hat{f}^{[0]}$  and  $r^{[0]}$
- Increase  $b$  by one, i.e.  $b \leftarrow b + 1$ , and compute the residuals  $\varepsilon^{[b-1]} = Y - r^{[b-1]}$
- From the sample  $\{\varepsilon_i^{[b-1]}, X_i\}_{i=1}^n$  compute  $\hat{T}_{\varepsilon^{[b-1]}}$  (defined in Eq. 4) and update the estimator of  $r$  as

$$r^{[b]} = r^{[b-1]} + \frac{\hat{T}_{\varepsilon^{[b-1]}}}{\hat{f}^{[0]}}$$

Repeat the above steps  $B$  times.

---

## 4 Test, boosting, and estimators

The boosting method has been originally proposed in the Machine Learning literature to handle the estimation bias. The application of  $L_2$  boosting algorithms to kernel estimation has been analyzed in several statistical papers, see ?, [Di Marzio & Taylor \(2008\)](#), [Park et al. \(2009\)](#), [Cornillon et al. \(2014\)](#). Essentially, the idea is to estimate an object of interest iteratively on different samples, and then build the final estimator by aggregating the estimates obtained from these samples. It is based on three main elements: (i) a starting value for the iteration; (ii) a base estimation procedure allowing to update the estimator at each iteration; (iii) the number of iterations to perform. The specific boosting algorithm we use in this paper is an  $L_2$ -boosting procedure. In this case, the base procedure for estimation consists in the minimization of a least squares criterion, see ? and [Di Marzio & Taylor \(2008\)](#). It is described in detail in Algorithm 1. Let  $r^{[0]}$  be an initial estimator of  $r$ . The boosting estimator after  $b$  iterations can be written as

$$r^{[b]} = r^{[b-1]} + \frac{\hat{T}_{\varepsilon^{[b-1]}}}{\hat{f}^{[0]}} \quad \hat{T}_{\varepsilon^{[b-1]}} = \frac{1}{nh^d} \sum_{i=1}^n \varepsilon_i^{[b-1]} \hat{t}_{n,i} K \left( \frac{W_i - \cdot}{h} \right) \quad (4)$$

for  $b = 1, \dots, B$ , where  $K$  is a kernel function,  $\varepsilon_i^{[b-1]} = Y_i - r^{[b-1]}(W_i)$ ,  $\hat{f}^{[0]}$  is an initial estimator of the density of  $W$ , and  $\hat{t}_{n,i}$  is a trimming factor used to handle a random denominator,

$$\hat{t}_{n,i} := 1' \hat{f}^{[0]}(W_i) \geq \tau_n' \quad t_{n,i} := 1' f(W_i) \geq \tau_n'.$$

The initial estimator  $r^{[0]}$  is called *weak learner*. The boosting updates in Algorithm 1 transform such weak learner into the *deep learner*  $r^{[B]}$  after  $B$  iterations. When  $\hat{f}^{[0]}$  and  $r^{[0]}$  are kernel estimators built with the same kernels and bandwidths as  $T$ , the boosting estimator of Eq. 4 is the same as in [Di Marzio & Taylor \(2008\)](#). Here, we are considering the initial estimators  $\hat{f}^{[0]}$  and  $r^{[0]}$  as generic, for the aim of generality of the theory and simplicity of proofs.

The statistic we use is

$$S_n := \int \sqrt{n} \mathbf{P}_{nB} \hat{\varepsilon}_n \varphi^2 \mu(dt), \text{ where } \hat{\varepsilon}_n = Y - r^{[B]}.$$

Our proofs also hold for a Kolgomorov-Smirnov transformation, and more in general for any continuous functional defined on  $A^\infty(\mathbb{T}) := \{g : \mathbb{T} \rightarrow \mathbb{R} \text{ s.t. } \sup_{t \in \mathbb{T}} |g(t)| < \infty\}$ .

## 5 Technical deftnitions and assumptions

Let  $W$  be the support of  $W$ . We assume that  $W$  is convex,  $d = \dim(W)$ , and define the set

$$W_n := \{w : f(w) > \frac{\tau_n}{2}\}. \quad (5)$$

For a vector of positive natural numbers  $k = (k_1, \dots, k_d)$ , define the differential operator

$$\partial^k := \frac{\partial^{k_1}}{\partial x_1} \dots \frac{\partial^{k_d}}{\partial x_d},$$

with  $k_+ = k_1 + \dots + k_d$ <sup>1</sup>. The class of smooth functions we consider is

$$\mathbf{C}(W) := \{g : W \rightarrow \mathbb{R} : \max_{k_+ \leq \frac{d^+}{2} + 1} \|\partial^k g\|_{\infty, W} \leq M\}, \quad (6)$$

with  $d^+$  the largest even number weakly smaller than  $d$ . If  $d$  is even then  $d^+ = d$ ; if  $d$  is odd then  $d^+ = d - 1$ .

**Observation:** Notice that  $\mathbf{C}(W) \subset \mathbf{C}_M^\alpha(W)$  for  $\alpha = \frac{d^+}{2} + 1$ , where  $\mathbf{C}_M^\alpha(W)$  is the class of functions defined at page 154 in [van der Vaart & Wellner \(1996\)](#).

Let  $\mathbf{K}_\lambda^{\bar{r}}$  be the class of functions  $(v_1, \dots, v_d) \mapsto \prod_{j=1}^d k(v_j)$  with  $k$  univariate kernel of order  $\bar{r}$  that is  $\lambda$  times continuously differentiable. Denote  $p_n = P(f(W) \leq \frac{\tau_n}{2})$  and  $\mathbf{C}_n = \mathbf{C}(W_n)$ .

**Assumption 1.**  $\{Y_i, W_i, X_i\}_{i=1}^n$  is a sequence of iid bounded random variables.

**Assumption 2.**  $r \in \mathbf{C}(W)$  and is  $\bar{r}$  times continuously differentiable with bounded derivatives.

**Assumption 3.**  $K \in \mathbf{K}_\lambda^{\bar{r}}$  for  $\lambda = \frac{d^+}{2} + 1$ .

**Assumption 4.**  $p_n n^{1/2} = o(1)$ ,  $\frac{p_n n^{1/4}}{h^{d\tau_n^B}} = o(1)$ ,  $\frac{h}{\tau_n} = o(1)$ , and for each  $n$  large enough  $W_n$  is a convex set.

---

<sup>1</sup>We are using the same notation as in [van der Vaart & Wellner, 1996](#)

**Assumption 5.**  $\frac{\log n}{n^{1/2} h^d \tau_n^{2B}} = o(1)$ ,  $nh^{4rB} = o(1)$ .

**Assumption 6.** (i) for  $b_n \in \{\hat{t}_n, t_n\}$ ,  $\|(\hat{r}^{[0]} - r)b_n\|_\infty = o(d_n)$ ,  $\|(\hat{f}^{[0]} - f)b_n\|_\infty = o(d_n)$ , with  $\frac{d_n}{\tau_n^B} = o(n^{-1/4})$ ; (ii)  $P(\hat{r}^{[0]} \in \mathcal{C}) \rightarrow 1$ ; (iii) for  $b = 1, \dots, B - 1$ :  $P\left(\frac{\hat{r}_{\varepsilon_b}}{f} \in \mathcal{C}_n\right) \rightarrow 1$ .

**Comments on the assumptions.** Assumptions 1–3 are standard. Assumption 4 is needed to take care of a random denominator by the trimming  $t_n$ . Assumption 5 establishes that the order of the kernel used in the boosting updates (i.e. for  $\hat{T}_{\varepsilon_b}$ ) must be decided in connection with the number of boosting iterations  $B$ . The appearance of  $\tau_n$  is due to the fact that at every boosting iteration we have to project the estimator  $\hat{r}^{[b-1]}$  obtained in the previous iteration, so we have to take care of a random denominator. Such a rate is introduced for theoretical reasons and is not relevant in the practical implementation of the test. If we ignore the presence of  $\tau_n$ , for  $B \geq 1$  Assumption 5 is avoiding undersmoothing, in the sense that the bandwidth minimizing the mean-squared error can be used for implementing the test. Assumption 6 is a high-level condition. In particular, Assumption 6(i) establishes the convergence rates of the weak learners. Such convergence rates are relatively standard in the literature and similar to e.g. Escanciano *et al.* (2014). Assumption 6(ii) imposes that the weak learner  $\hat{r}^{[0]}$  must belong to a class of functions sufficiently regular. This is also an assumption often made in the literature, see e.g. Escanciano *et al.* (2014), Mammen *et al.* (2012). Assumption 6(iii) is similar in spirit to Assumption 6(ii). It can be proved using the same arguments as in Appendix C, and in particular by the same arguments of Lemma C.6.

The following Proposition establishes the asymptotic behavior of the test statistic.

**Proposition 5.1.** Under Assumptions 1–6, if  $H_0$  holds,

$$S_n \xrightarrow{d} \int_0^1 \mathbf{G} \cdot \mu_\lambda(dt),$$

where  $\mathbf{G}$  is a Gaussian stochastic process taking values in  $A^\infty(\mathbf{G})$ ,  $\mathbf{G} := (y, x) \mapsto (y - r(w))$ ,  $\phi_t^\perp(x) : t \in \mathcal{T}$ , and  $\mathbf{G}$  is defined by the collection of covariances  $P\varepsilon^2 \phi_{t_1}^\perp \phi_{t_2}^\perp : t_1, t_2 \in \mathcal{T}$ .

## 6 The Bootstrap Test

Since the statistic is not asymptotically pivotal, for the computation of the critical value we propose a *Wild-Bootstrap* procedure that imposes the null hypothesis  $H_0$  when resampling the ob-

servations. The bootstrap DGP writes as

$$Y_i^* = \hat{r}^{[B]}(W_i) + \zeta_i \hat{\varepsilon}_{B,i} \text{ with } \hat{\varepsilon}_{B,i} = Y_i - \hat{r}^{[B]}(W_i), \quad (7)$$

where  $\{\zeta_i\}_{i=1}^n$  is a sequence of bootstrap weights with  $E\zeta = 0$  and  $E\zeta^2 = 1$ . The bootstrap version of  $S_n$  is

$$S_n^* = \int \sqrt{n} \mathbf{P}_{n,B} \hat{\varepsilon}_{B,t}^* \varphi_t^{\perp} d\mu(dt),$$

with

$$\hat{\varepsilon}_{B,i}^* = Y_i^* - \hat{r}^{*[B]}(W_i)$$

and

$$\hat{r}^{*[b]} = \hat{r}^{*[b-1]} + \frac{\hat{T}_{\hat{\varepsilon}_{B-1}^*}^*}{\hat{f}^{[0]}}, \quad \hat{T}_{\hat{\varepsilon}_{B-1}^*}^* = \frac{1}{nh^d} \sum_{i=1}^n \hat{\varepsilon}_{B-1,i}^* \hat{f}_{i,i} K \left( \frac{W_i - \cdot}{h} \right).$$

To show the validity of the bootstrap, we need to extend the regularity conditions in Assumption 6. Denote with  $P^*$  the probability measure that considers only  $\{\zeta_i\}_{i=1}^n$  as random and assumes as fixed the sample data  $\{Y_i, W_i, X_i\}_{i=1}^n$ . For a random variable  $Z$  measurable with respect to  $\{\zeta_i, Y_i, W_i, X_i\}_{i=1}^n$ , the notation  $Z_n = o_{P^*}(1)$  means that  $P^*(|Z| > \delta) \xrightarrow{P} 0$  for every  $\delta > 0$ .

**Assumption 7.** (i) Rates for the weak learners: for  $b_n \in \{t_n, \hat{t}_n\}$ ,  $\|(\hat{r}^{*[0]} - r)b_n\|_{\infty} = o_{P^*}(d_n)$ ,  $\|(\hat{f} - f)b_n\|_{\infty} = o_{P^*}(d_n)$ , with  $\frac{d_n}{t_n} = o(n^{-1/4})$ .

(ii) Regularity of the weak learner:  $P^*(\hat{r}^{*[0]} \in \mathcal{C}) \xrightarrow{P} 1$ ;

(iii) Regularity of the boosting updates: for  $b = 1, \dots, B-1$ :  $P_{\hat{f}}^{\hat{T}_{\hat{\varepsilon}_{B-1}^*}} \in \mathcal{C}_n \xrightarrow{P} 1$ ;

Denote with  $F_0$  the cdf of the real-valued random variable  $\int |G|^2 d\mu(dt)$  defined in Proposition 5.1, and let  $F_n$  denote the cdf of  $S_n$ . The validity of the bootstrap scheme with boosting iterations is shown by the following

**Proposition 6.1.** Under Assumptions 1-7 if  $H_0$  holds, uniformly over  $\mathcal{T}$ ,

$$\sqrt{n} \mathbf{P}_{n,B} \hat{\varepsilon}_{B,t}^* \hat{\varphi}_t = \sqrt{n} \mathbf{P}_{n,\zeta} \varepsilon \varphi_t^{\perp} + o_{P^*}(1).$$

Accordingly, for every continuity point of  $F_0$ ,

$$\mathbf{P}_{\zeta}(S_n \leq z) - F_n(z) \xrightarrow{P} 0.$$

Thanks to the previous proposition, the critical values can be simulated according to the DGP in Eq. 7. Repeating the bootstrap resampling a sufficiently large amount of times, say  $N^*$ , we obtain a collection of statistics  $\{S_{n,q}^*; q = 1, \dots, N^*\}$  whose distribution can be used to approximate the null distribution of  $S_n$ . The  $1 - \alpha$  quantile of such a distribution will approximate the true quantile of  $S_n$  and can be used as a critical value for running the test at the  $\alpha$  nominal level.

## 7 Simulations

In this section we provide evidence about the small-sample performance of our test. The DGP we consider is similar to the one provided in the counterexample of Section 2. The model writes as

$$Y = 1 + Z_1 + Z_3 + \gamma Z_2 + u,$$

where

$$u = g(Z_1, Z_2) + \eta \text{ and } g(Z_1, Z_2) = \begin{cases} Z_1^2 - \frac{1}{12} Z_2^2 & \text{if } Z_1 < 0 \\ 0 & \text{if } Z_1 \geq 0 \end{cases}$$

$Z_1$  and  $Z_3$  are resampled each from a uniform distribution with support  $[-0.5, +0.5]$  and are mutually independent. The residual  $\eta$  is independent from  $(Z_1, Z_2, Z_3)$ .  $Z_2$  is generated from a mixture-type distribution. In particular, denote with  $N(\mu, \sigma, a, b)$  a normal distribution that is truncated from below by  $a$ , from above by  $b$ , and has mean  $\mu$  and standard deviation  $\sigma$ . Then,

$$\begin{aligned} Z_2 &\sim N(0, .65, -0.5, 0.5) \text{ if } Z_1 < 0 \\ Z_2 &\sim N(0, .25, -0.5, 0.5) \text{ if } Z_1 \geq 0. \end{aligned}$$

Notice that the joint distribution of  $(Z_1, Z_2)$  provided here is equivalent to the one in Section 2, as long as  $\phi$  and  $\psi$  are set to  $N(0, .65, -0.5, 0.5)$  and  $N(0, .25, -0.5, 0.5)$ , respectively.

In terms of the notation used in the previous pages,  $W := (Z_1, Z_3)$  and  $X := (Z_2, Z_3)$ . According to the argumentation of Section 2, when  $\gamma = 0$  model  $M_1$  encompasses model  $M_2$ , but the other way around is not true. Moreover, if one wants to test the encompassing assumption, it is not sufficient to check the significance of  $Z_2$  in a nonparametric regression of  $Y$  onto  $(Z_1, Z_2, Z_3)$ : in this case  $Z_2$  would be significant, but  $M_1$  would encompass  $M_2$ .

For the implementation of the test, we employ Gaussian kernels of order 4, while the weighting function  $\phi$  is set to the complex exponential. We consider as a bandwidth rule  $h = Cn^{-1/6}$ , and we change the constant  $C$  over the set  $\{0.5, 1, 1.5, 2\}$ . The test is implemented with several iterations of the  $L_2$  boosting algorithm, ranging from 0 up to 5. For  $\gamma = 0$  we are under the null that model  $M_1$  encompasses model  $M_2$ , while for  $\gamma = 1$  we are under the alternative.

Table 1:  $C = 0.5$ 

| $n = 100$ |       |       |       |       |       |       | $n = 200$ |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|
| $H_0$     |       |       |       | $H_1$ |       |       | $H_0$     |       |       | H1    |       |       |
| $B$       | 0.01  | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   | 0.01      | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   |
| 0         | 0.024 | 0.056 | 0.13  | 0.32  | 0.572 | 0.686 | 0.016     | 0.058 | 0.124 | 0.68  | 0.862 | 0.926 |
| 1         | 0.02  | 0.062 | 0.134 | 0.324 | 0.584 | 0.688 | 0.018     | 0.064 | 0.138 | 0.682 | 0.86  | 0.922 |
| 2         | 0.02  | 0.07  | 0.142 | 0.324 | 0.576 | 0.68  | 0.02      | 0.068 | 0.152 | 0.678 | 0.858 | 0.928 |
| 3         | 0.02  | 0.074 | 0.146 | 0.334 | 0.582 | 0.682 | 0.02      | 0.07  | 0.15  | 0.682 | 0.862 | 0.922 |
| 4         | 0.02  | 0.074 | 0.144 | 0.336 | 0.588 | 0.688 | 0.02      | 0.072 | 0.158 | 0.678 | 0.862 | 0.92  |
| 5         | 0.022 | 0.076 | 0.15  | 0.34  | 0.59  | 0.692 | 0.022     | 0.074 | 0.16  | 0.68  | 0.862 | 0.918 |

Simulations based on 1000 replications.

The results of the simulations are reported in Table 1-4. As long as the test is implemented by undersmoothing, i.e. by choosing a small  $C$ , the test without boosting iterations behaves relatively well. The test with  $L_2$  boosting iterations appears to be competitive. Differently, as long as the constant  $C$  is increased, the test without iterations shows a very poor performance, while the test with  $L_2$  boosting iterations displays a relevant robustness and is able to control the size in a satisfactory way. These simulations show that the  $L_2$  boosting iterations are quite useful for controlling the empirical size of the test with respect to the choice of the bandwidth, especially when undersmoothing is avoided. Such a feature is attractive, as it is a diffused practice to select bandwidths by Cross-validation algorithms or other automatic procedures that do not guarantee undersmoothing. Moreover, there is no standard practice for selecting the bandwidth that guarantees undersmoothing. We also highlight the following feature. Compared to the nonparametric significance test in [Delgado & Manteiga \(2001\)](#), we cannot multiply the initial moment condition by a density that avoids a random denominator. Such a multiplication allows to stabilize the behavior of the test statistic and to make it relatively more robust with respect to the bandwidth choice. In our context, this is not possible, so the  $L_2$  boosting iterations appear to be an important tool to provide such robustness. The simulations also show a caveat: in the presence of a very low level of undersmoothing a large number of boosting iterations might be less able to control the size of the test, so it is advisable to avoid a very large number of  $B$ .

## 8 Conclusions

This paper provides a test to choose between two alternative models when each of them explains the same response variable. The choice between the two competing models is based on the encompassing principle, according to which a model  $M_1$  encompasses a model  $M_2$  if it can explain



Table 2:  $C = 1$ 

| $n = 100$ |       |       |       |       |       |       | $n = 200$ |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|
| $H_0$     |       |       |       | $H_1$ |       |       | $H_0$     |       |       | $H_1$ |       |       |
| $B$       | 0.01  | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   | 0.01      | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   |
| 0         | 0.022 | 0.112 | 0.196 | 0.334 | 0.594 | 0.708 | 0.03      | 0.108 | 0.246 | 0.704 | 0.906 | 0.936 |
| 1         | 0.016 | 0.058 | 0.116 | 0.308 | 0.548 | 0.674 | 0.014     | 0.056 | 0.116 | 0.662 | 0.864 | 0.912 |
| 2         | 0.016 | 0.054 | 0.106 | 0.316 | 0.57  | 0.668 | 0.016     | 0.052 | 0.112 | 0.66  | 0.868 | 0.916 |
| 3         | 0.018 | 0.054 | 0.108 | 0.314 | 0.566 | 0.668 | 0.018     | 0.05  | 0.114 | 0.668 | 0.868 | 0.912 |
| 4         | 0.018 | 0.056 | 0.112 | 0.310 | 0.566 | 0.666 | 0.018     | 0.048 | 0.116 | 0.666 | 0.872 | 0.914 |
| 5         | 0.018 | 0.054 | 0.11  | 0.304 | 0.568 | 0.668 | 0.018     | 0.052 | 0.116 | 0.67  | 0.87  | 0.914 |

Simulations based on 1000 replications.

Table 3:  $C = 1.5$ 

| $n = 100$ |       |       |       |       |       |       | $n = 200$ |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|
| $H_0$     |       |       |       | $H_1$ |       |       | $H_0$     |       |       | $H_1$ |       |       |
| $B$       | 0.01  | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   | 0.01      | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   |
| 0         | 0.102 | 0.368 | 0.52  | 0.456 | 0.698 | 0.812 | 0.288     | 0.628 | 0.77  | 0.868 | 0.964 | 0.978 |
| 1         | 0.032 | 0.136 | 0.256 | 0.35  | 0.608 | 0.734 | 0.04      | 0.174 | 0.338 | 0.728 | 0.914 | 0.954 |
| 2         | 0.02  | 0.082 | 0.16  | 0.31  | 0.556 | 0.684 | 0.014     | 0.074 | 0.15  | 0.686 | 0.884 | 0.922 |
| 3         | 0.016 | 0.072 | 0.13  | 0.302 | 0.544 | 0.664 | 0.012     | 0.064 | 0.12  | 0.67  | 0.87  | 0.916 |
| 4         | 0.012 | 0.062 | 0.118 | 0.3   | 0.538 | 0.668 | 0.012     | 0.054 | 0.118 | 0.67  | 0.868 | 0.916 |
| 5         | 0.012 | 0.062 | 0.112 | 0.3   | 0.552 | 0.668 | 0.012     | 0.05  | 0.116 | 0.664 | 0.872 | 0.918 |

Simulations based on 1000 replications.

Table 4:  $C = 2$ 

| $n = 100$ |       |       |       |       |       |       | $n = 200$ |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-----------|-------|-------|-------|-------|-------|
| $H_0$     |       |       |       | $H_1$ |       |       | $H_0$     |       |       | $H_1$ |       |       |
| $B$       | 0.01  | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   | 0.01      | 0.05  | 0.1   | 0.01  | 0.05  | 0.1   |
| 0         | 0.206 | 0.486 | 0.614 | 0.512 | 0.766 | 0.854 | 0.574     | 0.804 | 0.89  | 0.914 | 0.98  | 0.986 |
| 1         | 0.094 | 0.338 | 0.498 | 0.44  | 0.684 | 0.804 | 0.238     | 0.584 | 0.73  | 0.86  | 0.962 | 0.978 |
| 2         | 0.044 | 0.184 | 0.338 | 0.39  | 0.642 | 0.75  | 0.072     | 0.304 | 0.488 | 0.768 | 0.93  | 0.966 |
| 3         | 0.024 | 0.132 | 0.226 | 0.338 | 0.598 | 0.72  | 0.032     | 0.14  | 0.296 | 0.722 | 0.904 | 0.95  |
| 4         | 0.018 | 0.098 | 0.17  | 0.314 | 0.556 | 0.69  | 0.016     | 0.086 | 0.182 | 0.688 | 0.888 | 0.924 |
| 5         | 0.02  | 0.078 | 0.146 | 0.296 | 0.548 | 0.678 | 0.014     | 0.068 | 0.138 | 0.676 | 0.878 | 0.918 |

Simulations based on 1000 replications.

the results of  $M_2$ . The test we propose is based on a continuum of moments and is fully non-parametric. Hence, it does not rely on neither specific functional forms nor on fixed values of the explanatory variables. We implement the test by an  $L_2$  boosting algorithm which allows us to avoid undersmoothing. We propose a wild-bootstrap procedure for the computation of the critical values and prove its validity with boosting iterations. A Monte Carlo simulation study shows the attractiveness of boosting. Such an algorithm allows the test to be relatively robust to the bandwidth choice, especially when undersmoothing does not hold. This is indeed an attractive feature, as existing methods for bandwidth selection, like Cross-Validation methods, do not allow to select a bandwidth respecting the undersmoothing conditions.

## References

- Bierens, Herman J. 1982. Consistent model specification tests. *Journal of Econometrics*, **20**(1), 105–134. [3](#)
- Bierens, Herman J., & Ploberger, Werner. 1997. Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica*, **65**(5), 1129–1152. [3](#)
- Bontemps, Christophe, & Mizon, Grayham E. 2008. Encompassing: Concepts and Implementation\*. *Oxford Bulletin of Economics and Statistics*, **70**(s1), 721–750. [1](#)
- Bontemps, Christophe, Florens, Jean-Pierre, & Richard, Jean-Francois. 2008. Parametric and Non-parametric Encompassing Procedures\*. *Oxford Bulletin of Economics and Statistics*, **70**(s1), 751–780. [1](#)
- Cornillon, Pierre-Andre, Hengartner, N. W., & Matzner Lober, E. 2014. Recursive bias estimation for multivariate regression smoothers. *ESAIM: Probability and Statistics*, **18**, 483–502. [4](#)
- Delgado, Miguel A., & Manteiga, Wenceslao Gonzalez. 2001. Significance testing in non-parametric regression based on the bootstrap. *The Annals of Statistics*, **29**(5), 1469–1507. [3](#), [7](#)
- Di Marzio, Marco, & Taylor, Charles C. 2008. On boosting kernel regression. *Journal of Statistical Planning and Inference*, **138**(8), 2483–2498. [4](#), [4](#)
- Escanciano, Juan Carlos, Jacho-Chávez, David T., & Lewbel, Arthur. 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics*, **178**(P3), 426–443. [3](#), [5](#)

- Florens, Jean-Pierre, Hendry, David F., & Richard, Jean-Francois. 1996. Encompassing and Specificity. *Econometric Theory*, **12**(4), 620–656. [1](#)
- Gourieroux, Christian, & Monfort, Alain. 1995. Testing, Encompassing, and Simulating Dynamic Econometric Models. *Econometric Theory*, **11**(2), 195–228. [1](#), [2.1](#)
- Gourieroux, Christian, Monfort, Alain, & Trognon, Alain. 1983. Testing nested or non-nested hypotheses. *Journal of Econometrics*, **21**(1), 83–115. [1](#)
- Hendry, David, & Richard, Jean-Francois. 1987. Recent developments in the theory of encompassing. *Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), CORE Discussion Papers*, Jan. [1](#)
- Li, Qi, & Racine, Jeffrey Scott. 2006. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press. [C](#)
- Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2012. Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics*, **40**(2), 1132–1170. [5](#)
- Mizon, Grayham E., & Richard, Jean-Francois. 1986. The Encompassing Principle and its Application to Testing Non-Nested Hypotheses. *Econometrica*, **54**(3), 657–678. [1](#)
- Park, B. U., Lee, Y. K., & Ha, S. 2009. L2 boosting in kernel regression. *Bernoulli*, **15**(3), 599–613. [4](#)
- Stinchcombe, Maxwell B., & White, Halbert. 1998. Consistent Specification Testing With Nuisance Parameters Present Only Under The Alternative. *Econometric Theory*, **14**(03), 295–325. [3](#)
- van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge University Press. [A](#), [C](#), [C](#)
- van der Vaart, A. W., & Wellner, Jon. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media. [5](#), [1](#), [B](#), [C](#)

## A Asymptotic Expansion

Before proving the main results, we introduce a class of functions that will be often used in the technical proofs.

**Condition (CL).**  $\Psi := \{x \mapsto \psi_t(x) : t \in \mathbf{T}\}$  is a collection of uniformly bounded Lipschitz functions in  $t$ , i.e.  $\|\psi_{t_1} - \psi_{t_2}\|_\infty \leq C|t_1 - t_2|$  for all  $t_1, t_2 \in \mathbf{T}$ . We define  $\phi_t(W) := \mathbf{E}\{\psi_t(X) | W\}$ .

We start with a technical lemma that will be needed in the derivation of the Brahadur representation of the empirical process at the basis of  $S_n$ .

**Lemma A.1. (Stochastic Equicontinuity Results)** *Let Assumption 6 hold, and let  $\Psi$  and  $\phi_t$  be as in Condition [CL](#). Then, for  $b = 0, \dots, B$  uniformly in  $t \in \mathbb{T}$ ,*

- (i)  $\sqrt{n}(\mathbf{P}_n - P) \frac{\hat{T}_{\varepsilon b}}{f} t_n \psi_t = o_p(1)$ ;
- (ii)  $\sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon b}}{f} t_n \psi_t = \sqrt{n} \mathbf{P}_n \varepsilon \phi_t + \sqrt{n} \mathbf{P}_n (r - \hat{r}^{[b]}) t_n \phi_t + h^r \sqrt{n} \mathbf{P}_n (Y - \hat{r}^{[b]}) t_n \phi_{n,t}^{(1)} + o_p(1)$ , where  $\phi_{n,t}$  satisfies Condition [CL](#);
- (iii)  $\sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{[b]} - r) t_n \psi_t = o_p(1)$ , and the same result holds by replacing  $t_n$  with  $\hat{t}_n$ .

*Proof.* For Point (i), by Lemma [C.6](#), Assumption 4, 5, and 6, we obtain

$$\sup_{t \in \mathbb{T}} \left\| \frac{\hat{T}_{\varepsilon b}}{f} t_n \psi_t \right\|_{\infty} = o_p(1).$$

By the above display, Assumption 6(iii), and Lemma [C.1](#), we can apply Lemma [C.3](#) to obtain Point (i).

For Point (ii), by the Law of Iterated Expectations, the definition of  $T$ ,  $\hat{\varepsilon}_b$  and Assumption 4,

$$\begin{aligned} \sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon b}}{f} t_n \psi_t &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_i^{[b]}) \int_{\hat{t}_{n,i}}^{\hat{t}_{n,i} + h} K(v) \phi_t(W_i + vh) t_n(W_i + vh) dv = \\ &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_i^{[b]}) \int_{\hat{t}_{n,i}}^{\hat{t}_{n,i} + h} K(v) \phi_t(W_i + vh) dv = \\ &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_i^{[b]}) \hat{t}_{n,i} \phi_{n,t,i} + h^r \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_i^{[b]}) \hat{t}_{n,i} \phi_{n,t,i}^{(1)}, \end{aligned}$$

where we have used the usual change of variable for the first equality and an  $r$ -th order Taylor expansion of  $\phi_t$  for the third one. We can add and subtract  $r$  in the first addendum of the above display. Then, by using Lemma [C.2](#) and [C.6](#), we can replace the trimming  $\hat{t}_n$  with  $t_n$  in the resulting expression. The result of Point (ii) therefore follows by noticing that  $\sqrt{n} \mathbf{P}_n \varepsilon t_n \phi_t = \sqrt{n} \mathbf{P}_n \varepsilon \phi_t + o_p(1)$  uniformly over  $\mathbb{T}$ , by Assumption 4.

For Point (iii), by using the recursive structure in Eq. [4](#),

$$\sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{[b]} - r) t_n \psi_t = \sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{[0]} - r) t_n \psi_t + \sum_{s=0}^{b-1} \sqrt{n}(\mathbf{P}_n - P) \frac{\hat{T}_{\varepsilon s}}{f \cdot f^{[0]}} t_n \psi_t.$$

Assumption 6(i)-(ii) and Lemma [C.1](#) combined with Lemma [C.3](#) ensure that the first term on the RHS of the above display is  $o_p(1)$  uniformly over  $\mathbb{T}$ . For a generic addendum of the second term on the RHS,

$$\sqrt{n}(\mathbf{P}_n - P) \frac{\hat{T}_{\varepsilon s}}{f \cdot f^{[0]}} t_n \psi_t = \sqrt{n}(\mathbf{P}_n - P) \frac{\hat{T}_{\varepsilon s}}{f} t_n \psi_t + \sqrt{n}(\mathbf{P}_n - P) \hat{T}_{\varepsilon s} \sum_{f \cdot f^{[0]}} \sum_{f \cdot f^{[0]}} \psi_t.$$

By Point 1, Lemma [C.6](#), and Assumption 5-6, the above expression is  $o_P(1)$  uniformly in  $t$ . By this and the two previous display, we conclude for Point (iii).  $\square$

**Lemma A.2. (First expansion)** Under Assumption 1-6, uniformly over  $\mathbb{T}$ ,

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{t}_n \varphi_t = \sqrt{n} \mathbf{P}_n \varepsilon \varphi_t^\perp + h^r \sqrt{n} \mathbf{P}_n (r^{\hat{[B-1]}} - r) t_n \beta_{n,t} + o_P(1),$$

where  $\beta_{n,t}$  satisfies Condition [CL](#).

*Proof.* Since  $\hat{\varepsilon}_B = Y - r^{\hat{[B]}}$ , using Lemma [C.2](#) and [C.6](#), together with Assumption 4,

$$\begin{aligned} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{t}_n \varphi_t &= \sqrt{n} \mathbf{P}_n \varepsilon \hat{t}_n \varphi_t + \sqrt{n} \mathbf{P}_n (r - r^{\hat{[B]}}) \hat{t}_n \varphi_t = \\ &= \sqrt{n} \mathbf{P}_n \varepsilon \varphi_t + \sqrt{n} \mathbf{P}_n (r - r^{\hat{[B]}}) t_n \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathbb{T}$ . From the recursive structure in Eq. [4](#),

$$\begin{aligned} \sqrt{n} \mathbf{P}_n (\hat{r}^{[B]} - r) t_n \varphi_t &= \sqrt{n} \mathbf{P}_n (\hat{r}^{[B-1]} - r) t_n \varphi_t + \sqrt{n} \mathbf{P}_n \frac{\hat{r}_{\varepsilon_{B-1}}}{f} t_n \varphi_t + \sqrt{n} \mathbf{P}_n \frac{\hat{r}_{\varepsilon_{B-1}}}{f} t_n \varphi_t \sum_{f \cdot f^{[0]}} \sum_{f \cdot f^{[0]}} \varphi_t = \\ &= \sqrt{n} \mathbf{P}_n (\hat{r}^{[B-1]} - r) t_n \varphi_t + \sqrt{n} \mathbf{P}_n \frac{\hat{r}_{\varepsilon_{B-1}}}{f} t_n \varphi_t + o_P(1), \end{aligned} \quad (8)$$

where in the last equality we have used Lemma [C.6](#) and Assumptions 4-6. By Lemma [A.1](#) (i) and (ii),

$$\sqrt{n} \mathbf{P}_n \frac{\hat{r}_{\varepsilon_{B-1}}}{f} t_n \varphi_t = \sqrt{n} \mathbf{P}_n \varepsilon l_t + \sqrt{n} \mathbf{P}_n (r - \hat{r}^{[B-1]}) t_n l_t + h^r \sqrt{n} \mathbf{P}_n (Y - r^{\hat{[B-1]}}) t_n l_{n,t}^{(1)} + o_P(1).$$

Now, by Assumption 4 and Lemma [C.3](#), uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} h^r \sqrt{n} \mathbf{P}_n (Y - \hat{r}^{[B-1]}) t_n l_{n,t}^{(1)} &= h^r \sqrt{n} \mathbf{P}_n \varepsilon l_{n,t}^{(1)} + h^r \sqrt{n} \mathbf{P}_n (r - r^{\hat{[B-1]}}) t_n l_{n,t}^{(1)} + o_P(1) = \\ &= h^r \sqrt{n} \mathbf{P}_n (r - r^{\hat{[B-1]}}) t_n l_{n,t}^{(1)} + o_P(1). \end{aligned}$$

By putting together the previous four displays and simplifying,

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{t}_n \varphi_t = \sqrt{n} \mathbf{P}_n \varepsilon \varphi_t^\perp - h^r \sqrt{n} \mathbf{P}_n (r^{\hat{[B-1]}} - r) t_n \beta_{n,t} - \sqrt{n} \mathbf{P}_n (r^{\hat{[B-1]}} - r) t_n \varphi_t^\perp + o_P(1).$$

Finally, notice that  $P(r^{\hat{[B-1]}} - r) t_n \varphi_t^\perp = 0$ , so the result of the Lemma follows by applying Lemma [A.1](#) (iii) to the third term on the RHS of the above expression.  $\square$

**Lemma A.3. (Negligibility of the boosting iterations)** Under Assumptions 1-6, uniformly in  $t \in \mathbb{T}$ ,

$$h^r \sqrt{n} \mathbf{P}_n(r^{[B-1]} - r) t_n \tilde{\beta}_{n,t} = o_P(1).$$

*Proof.* We first show the following recursive structure:

$$h^{rs} \sqrt{n} \mathbf{P}_n(r^{[B-s]} - r) t_n \tilde{\beta}_{n,t} = -h^{r(s+1)} \sqrt{n} \mathbf{P}_n(r^{[B-(s+1)]} - r) t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1) \quad (9)$$

for  $s = 1, \dots, B-1$

uniformly over  $\mathbb{T}$ . To this end, from Eq. 4 we obtain

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n(\hat{r}^{[B-s]} - r) t_n \tilde{\beta}_{n,t} &= h^{rs} \sqrt{n} \mathbf{P}_n(\hat{r}^{[B-(s+1)]} - r) t_n \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\varepsilon}_{B-(s+1)}}}{f^{[0]}} t_n \tilde{\beta}_{n,t} = \\ &= h^{rs} \sqrt{n} \mathbf{P}_n(r^{[B-(s+1)]} - r) t_n \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\varepsilon}_{B-(s+1)}}}{f} t_n \tilde{\beta}_{n,t} + o_P(1) \end{aligned} \quad (10)$$

uniformly over  $\mathbb{T}$ , where the last equality follows by proceeding similarly as in Eq 8. By applying Lemma A.1 to the second term on the RHS of the previous display,

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\varepsilon}_{B-(s+1)}}}{f} t_n \tilde{\beta}_{n,t} &= h^{rs} \sqrt{n} \mathbf{P}_n \varepsilon \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n(r - r^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t} + \\ &+ h^{r(s+1)} \sqrt{n} \mathbf{P}_n(Y - \hat{r}^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1). \end{aligned}$$

By Lemma C.3,  $\sqrt{n} \mathbf{P}_n \varepsilon h^r \tilde{\beta}_{n,t} = o_P(1)$  uniformly in  $t \in \mathbb{T}$ . Similarly,  $h^{r(s+1)} \sqrt{n} \mathbf{P}_n(Y - \hat{r}^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t}^{(1)} = h^{r(s+1)} \sqrt{n} \mathbf{P}_n(r - \hat{r}^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1)$  uniformly over  $\mathbb{T}$ . By plugging these approximations in the above display,

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\varepsilon}_{B-(s+1)}}}{f} t_n \tilde{\beta}_{n,t} &= \\ h^{rs} \sqrt{n} \mathbf{P}_n(r - \hat{r}^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t} + h^{r(s+1)} \sqrt{n} \mathbf{P}_n(r - \hat{r}^{[B-(s+1)]}) t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1). \end{aligned} \quad (11)$$

The recursive structure of Eq. 9 follows from Eq. 10 and 11.

Since Eq. 9 holds for any  $s = 1, \dots, B-1$ , if we proceed by a simple induction we obtain

$$h^r \sqrt{n} \mathbf{P}_n(r^{[B-1]} - r) t_n \tilde{\beta}_{n,t} = (-1)^{B-1} h^{rB} \sqrt{n} \mathbf{P}_n(r^{[0]} - r) t_n \tilde{\beta}_{n,t} + o_P(1)$$

uniformly over  $\mathbb{T}$ . By Assumption 6  $\|(\hat{r}^{[0]} - r) t_n\|_\infty = o_P(n^{-1/4})$  and by Assumption 5  $nh^{4rB} = o(1)$ . By these rates and the above display, we conclude.  $\square$

The above lemma shows that, under Assumptions 1-6, the boosting iterations do not have an

impact on the Bahadur expansion of the empirical process  $\sqrt{n}P_n\hat{\varepsilon}_B\hat{t}_n\varphi_t$  at the basis of the statistic  $S_n$ . This results and the asymptotic distribution of  $S_n$  are reported in the following:

**Corollary A.4. (Proof of Proposition 5.1.)** Under Assumption 1-6,

$$\sqrt{n}P_n\hat{\varepsilon}_B\hat{t}_n\varphi_t = \sqrt{n}P_n\varepsilon\varphi_t^\perp + o_P(1)$$

uniformly over  $\mathcal{T}$ . Accordingly,

$$S_n = \int \sqrt{n}P_n\hat{\varepsilon}_B\hat{t}_n\varphi_t^2\mu(dt) + \int G^2\mu(dt),$$

where  $G$  is a Gaussian stochastic process taking values in  $A^\infty(G)$ , with  $G := (y, x) \mapsto (y - r(w)) \cdot \varphi_t^\perp(x) : t \in \mathcal{T}$ , and defined by the collection of covariances  $P\varepsilon^2\phi_{t_1}\phi_{t_2} : t_1, t_2 \in \mathcal{T}$ .

*Proof.* The first result is an immediate consequence of Lemma A.2 and A.3. For the second result, we can proceed as in Example 19.7 in [van der Vaart \(1998\)](#) by the compactness of  $\mathcal{T}$  and the boundedness of the random variables involved, to obtain that  $N(\delta, G, \|\cdot\|_\infty) \leq N(C\delta, \mathcal{T}, \|\cdot\|)$ , with  $\int_0^1 \sqrt{\log N(\delta, \mathcal{T}, \|\cdot\|)} d\delta < \infty$ . By Theorem 19.14 in [van der Vaart \(1998\)](#),  $\sqrt{n}P_n\varepsilon\varphi_t^\perp \rightarrow G$ . Hence, conclude by Theorem 18.11(i) in [van der Vaart \(1998\)](#).  $\square$

**Remark A.5.** The expansion  $\sqrt{n}P_n\hat{\varepsilon}_B\hat{t}_n\varphi_t = \sqrt{n}P_n\varepsilon\varphi_t^\perp + o_P(1)$  proved in the above proposition also holds for any weighting function  $\psi_t$  such that  $\|\psi_{t_1} - \psi_{t_2}\|_\infty \leq C\|t_1 - t_2\|$  for all  $t_1, t_2 \in \mathcal{T}$ .

## B Bootstrap Expansion

In this section we prove the validity of the bootstrap test with boosting. The general structure of the proof is similar to the one of the previous section.

**Lemma B.1. (Stochastic Equicontinuity for the Bootstrap)** Let  $\Psi$  and  $\phi_t$  be as in Condition [CL](#). Then, for  $b = 0, \dots, B$  uniformly in  $t \in \mathcal{T}$ ,

- (i)  $\sqrt{n}P_n\hat{r}^{*[b]}(\hat{t}_n - t_n)\psi_t = o_P(1)$ ;
- (ii)  $\sqrt{n}(P_n - P)\frac{\hat{T}_{\varepsilon_b}^*}{f}\psi_t = o_P(1)$ ;
- (iii)  $\sqrt{n}P\frac{\hat{T}_{\varepsilon_b}^*}{f}\psi_t = \sqrt{n}P_n\varepsilon\xi\phi_t + \sqrt{n}P_n(r^{[B]} - r^{*[b]})t_n\phi_t + h^r\sqrt{n}P_n\varepsilon_b^*t_n\phi_{n,t}^{(1)}(W) + o_P(1)$ , where  $\phi_{n,t}^{(1)}$  satisfies Condition [CL](#)

$$(iv) \sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{*[b]} - r^{[b]})t_n\psi_t = o_P(1).$$

*Proof.* For Point (i), notice that since  $\hat{t}_n - t_n = (\hat{t}_n + t_n)(\hat{t}_n - t_n)$ ,

$$\sqrt{n}\mathbf{P}_n\hat{r}^{*[b]}(\hat{t}_n - t_n)\psi_t = \sqrt{n}\mathbf{P}_n(\hat{r}^{*[b]} - r^{[0]})(\hat{t}_n + t_n)(\hat{t}_n - t_n)\psi_t + \sqrt{n}\mathbf{P}_n r^{[0]}(\hat{t}_n - t_n)\psi_t.$$

By Assumption 6 and Lemma C.2, the second term of the previous display is  $o_P(1)$  uniformly over  $\mathbf{T}$ . For the first term, from Lemma C.7  $\|(\hat{r}^{*[b]} - r^{[0]})(\hat{t}_n + t_n)\|_\infty = o_P(1)$ , and from Lemma C.2  $\sqrt{n}\mathbf{P}_n|\hat{t}_n - t_n| = o_P(1)$ , so we conclude for the result in Point (i).

The proof of Point (ii) uses exactly the same arguments as the proof of Lemma A.1, Point (i), together with Assumption 7, Lemma C.4, and Lemma C.7.

For Point (iii), by the law of iterated expectations, the definition of  $T_{\varepsilon_b^*}^*$  and the classical change of variable,

$$\begin{aligned} \sqrt{n}P_{\frac{\hat{T}_{\varepsilon_b^*}}{f}}\psi_t &= \frac{1}{n} \sum_{i=1}^n \varepsilon_{b,i}^* \int K(u)t_n(W_i + uh)\phi_t(W_i + uh)du = \\ &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(u)\phi_t(W_i + uh)du + o_P(1), \end{aligned}$$

where for the second equality we have used  $\|(\hat{r}^{*[b]} - r^{[0]})\hat{t}_n\|_\infty = o_P(1)$  which is implied by Lemma C.7,  $\|(\hat{r}^{[b]} - r)\hat{t}_n\|_\infty = o_P(1)$  which is implied by Lemma C.6,  $h/\tau_n = o(1)$ , and Markov's inequality. By an  $r$ th order Taylor expansion,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \varepsilon_{b,i}^* \int K(u)\phi_t(W_i + uh)du &= \sqrt{n}\mathbf{P}_n\hat{\varepsilon}_b^*\phi_t(W_i) + h^r \sqrt{n}\mathbf{P}_n\hat{\varepsilon}_b^*t_n\phi_{t,t}(W) = \\ &= \sqrt{n}\mathbf{P}_n\hat{\varepsilon}_b^*t_n\phi_t(W) + h^r \sqrt{n}\mathbf{P}_n\hat{\varepsilon}_b^*t_n\phi_{n,t}(W) + o_P(1), \end{aligned}$$

with the second equality ensured by from Point (i) and Lemma C.6. From the definition of  $\hat{\varepsilon}_b^*$ , the expansion in Corollary A.4 (see also Remark A.5), and since  $E\zeta = 0$ ,

$$\begin{aligned} \sqrt{n}\mathbf{P}_n\hat{\varepsilon}_b^*t_n\phi_t(W) &= \sqrt{n}\mathbf{P}_n(Y - \hat{r}^{[B]})t_n\zeta\phi_t + \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[b]})t_n\phi_t = \\ &= \sqrt{n}\mathbf{P}_n\varepsilon\zeta\phi_t + \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[b]})t_n\phi_t + o_P(1). \end{aligned}$$

Since  $o_P(1) = o_P(1)$ , we can conclude for Point (iii) by putting together the previous three displays.

For Point (iv), notice first that by the rates in Assumption 6, Lemma C.6, and Lemma C.7,

$$\begin{aligned} \sqrt{n}(\mathbf{P}_n - P)\frac{\hat{T}_{\varepsilon_b^*}}{f}t_n\psi_t &= \sqrt{n}(\mathbf{P}_n - P)\frac{\hat{T}_{\varepsilon_b^*}}{f}t_n\psi + o_P(1) \text{ and} \\ \sqrt{n}(\mathbf{P}_n - P)\frac{\hat{T}_{\varepsilon_b^*}}{f}t_n\psi_t &= \sqrt{n}(\mathbf{P}_n - P)\frac{\hat{T}_{\varepsilon_b^*}}{f}t_n\psi + o_P(1). \end{aligned}$$



The two leading terms on the RHS of the above expressions are respectively  $o_P(1)$  and  $o_{P^*}(1)$ , from Point (ii) of the present Lemma and Lemma A.1. Hence, by the recursive structures of  $\hat{r}^{*[b]}$  and  $\hat{r}^{[b]}$ ,

$$\sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{*[b]} - \hat{r}^{[b]})t_n\psi_t = \sqrt{n}(\mathbf{P}_n - P)(\hat{r}^{*[0]} - \hat{r}^{[0]})t_n\psi_t + o_{P^*}(1) = o_{P^*}(1),$$

where the second equality follows from Assumption 6 and Lemma C.4. So, Point (iv) is also proved.  $\square$

**Lemma B.2. (A First Bootstrap Expansion)** Under Assumptions 1-6 uniformly in  $t \in \mathcal{T}$ ,

$$\sqrt{n}\mathbf{P}_n\hat{\varepsilon}_B^*t_n\varphi_t = \sqrt{n}\mathbf{P}_n\zeta\varepsilon\varphi_t^\perp - h^r\sqrt{n}\mathbf{P}_n\hat{\varepsilon}_B^*t_n\beta_{n,t} + o_{P^*}(1).$$

*Proof.* From the definitions of  $\hat{\varepsilon}_B^*$  and  $Y^*$ ,

$$\sqrt{n}\mathbf{P}_n\hat{\varepsilon}_B^*t_n\varphi_t = \sqrt{n}\mathbf{P}_n(Y - \hat{r}^{[B]})\hat{t}_n\zeta\varphi_t + \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[B]})\hat{t}_n\varphi_t. \quad (12)$$

Using the expansion in Corollary A.4 and Remark A.5, since  $E\zeta = 0$ ,

$$\sqrt{n}\mathbf{P}_n(Y - \hat{r}^{[B]})\hat{t}_n\zeta\varphi_t = \sqrt{n}\mathbf{P}_n\varepsilon\zeta\varphi_t + o_P(1) \text{ uniformly over } \mathcal{T}. \quad (13)$$

From Lemma C.6, C.2, and B.1, uniformly in  $t \in \mathcal{T}$ ,

$$\begin{aligned} \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[B]})\hat{t}_n\varphi_t &= \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[B]})t_n\varphi_t + o_{P^*}(1) = \\ \sqrt{n}\mathbf{P}_n(\hat{r}^{[B-1]} - \hat{r}^{*[B-1]})t_n\varphi_t + \sqrt{n}\mathbf{P}_n\frac{\hat{T}_{\varepsilon_{B-1}}^*}{f}t_n\varphi_t - \sqrt{n}\mathbf{P}_n\frac{\hat{T}_{\varepsilon_{B-1}}^*}{f}t_n\varphi_t + o_{P^*}(1). \end{aligned}$$

From the convergence rates in Assumption 6 and Lemma C.7, we can replace  $\hat{f}$  with  $f$  in the last term on the RHS of the previous display. Then, from Lemma B.1, uniformly over  $\mathcal{T}$ ,

$$\begin{aligned} \sqrt{n}\mathbf{P}_n\frac{\hat{T}_{\varepsilon_{B-1}}^*}{f}t_n\varphi_t &= \sqrt{n}\mathbf{P}_n\frac{\hat{T}_{\varepsilon_{B-1}}^*}{f}t_n\varphi_t + o_P(1) = \\ \sqrt{n}\mathbf{P}_n\varepsilon\zeta\varphi_t + \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[B-1]})t_n\varphi_t + h^r\sqrt{n}\mathbf{P}_n\hat{\varepsilon}_B^*t_n\beta_{n,t}(W) + o_{P^*}(1). \end{aligned}$$

Putting together the previous two displays,

$$\begin{aligned} \sqrt{n}\mathbf{P}_n(\hat{r}^{[B]} - \hat{r}^{*[B]})\hat{t}_n\varphi_t &= -\sqrt{n}\mathbf{P}_n\zeta\varepsilon\varphi_t + \sqrt{n}\mathbf{P}_n(\hat{r}^{[B-1]} - \hat{r}^{*[B-1]})t_n\varphi_t + \\ \sqrt{n}\mathbf{P}_n\frac{\hat{T}_{\varepsilon_{B-1}}^*}{f}t_n\varphi_t - h^r\sqrt{n}\mathbf{P}_n\hat{\varepsilon}_B^*t_n\beta_{n,t}(W) + o_{P^*}(1). \end{aligned} \quad (14)$$

From the rates in Assumption 6 and Lemma C.7 we can replace  $\hat{f}$  with  $f$  in the third term on the RHS of the above expression. Then, by Lemma B.1, we obtain that uniformly over  $\mathcal{T}$ ,

$$\sqrt{n} \mathbf{P}_n(r^{[B-1]} - \hat{r}^{*[B-1]}) t_n \varphi_t^\perp = o_P^*(1) \text{ and } \sqrt{n} \mathbf{P}_n \frac{\hat{\varepsilon}_{B-1}^*}{\hat{f}} t_n \varphi_t^\perp = o_P^*(1). \quad (15)$$

From Eq. 12, 13, 14, and 15 we conclude.  $\square$

**Lemma B.3. (Negligibility of the Reminder Term in the Bootstrap Expansion)** Under Assumption 1-6, uniformly over  $\mathcal{T}$ ,

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \beta_{n,t} = o_P^*(1)$$

*Proof.* We first obtain the following recursive structure:

$$h^{rS} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-S}^* t_n \beta_{n,t} = -h^{r(S+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(S+1)}^* t_n \beta_{n,t}^{(1)} + o_P^*(1) \text{ uniformly over } \mathcal{T}. \quad (16)$$

for  $S < B$ . So, consider  $S < B$ . By definition of  $\hat{\varepsilon}_{B-S}^*$ ,

$$h^{rS} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-S}^* t_n \beta_{n,t} = h^{rS} \sqrt{n} \mathbf{P}_n (Y - \hat{r}^{[B]}) t_n \xi \beta_{n,t} + h^{rS} \sqrt{n} \mathbf{P}_n (\hat{r}^{[B]} - \hat{r}^{*[B-S]}) t_n \beta_{n,t}. \quad (17)$$

By the expansion in Corollary A.4, Remark A.5, and since  $E\xi = 0$ , uniformly in  $t \in \mathcal{T}$ ,

$$h^{rS} \sqrt{n} \mathbf{P}_n (Y - \hat{r}^{[B]}) t_n \xi \beta_{n,t} = h^{rS} \sqrt{n} \mathbf{P}_n \varepsilon \xi \beta_{n,t} + o_P(1) = o_P(1). \quad (18)$$

Using the recursive structure of  $\hat{r}^{[b]}$ ,

$$\begin{aligned} h^{rS} \sqrt{n} \mathbf{P}_n (\hat{r}^{[B]} - \hat{r}^{*[B-S]}) t_n \beta_{n,t} = \\ h^{rS} \sqrt{n} \mathbf{P}_n (\hat{r}^{[B]} - \hat{r}^{*[B-(S+1)]}) t_n \beta_{n,t} - h^{rS} \sqrt{n} \mathbf{P}_n \frac{\hat{\varepsilon}_{B-(S+1)}^*}{\hat{f}} t_n \beta_{n,t}. \end{aligned} \quad (19)$$

Consider now the second term on the RHS of the previous display. The rates in Assumption 6 and Lemma C.7 ensure that we can replace  $\hat{f}$  with  $f$ . Then, Points (ii) and (iii) of Lemma B.1 deliver

$$\begin{aligned} h^{rS} \sqrt{n} \mathbf{P}_n \frac{\hat{\varepsilon}_{B-(S+1)}^*}{\hat{f}} t_n \beta_{n,t} = h^{rS} \sqrt{n} \mathbf{P}_n \xi \varepsilon \beta_{n,t} + \\ h^{rS} \sqrt{n} \mathbf{P}_n (\hat{r}^{[B]} - \hat{r}^{[B-(S+1)]}) t_n \beta_{n,t} + h^{r(S+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(S+1)}^* t_n \beta_{n,t}^{(1)} + o_P^*(1) \end{aligned} \quad (20)$$

uniformly over  $\mathcal{T}$ . Putting together Eq. 17, 18, 19, and 20 yields Eq. 16.

Since Eq. 16 holds for  $S = 1, \dots, B - 1$ , we can proceed to a simple induction to obtain

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \beta_{n,t} = (-1)^{B-1} h^B \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_0^* t_n \tilde{\beta}_{n,t}^{(1)} + o_P^*(1) \text{ uniformly over } \mathbb{T}.$$

Recall that  $\hat{\varepsilon}_0 = \zeta(Y - r^{[B]}) + r^{[B]} - r^{*[0]}$ . The expansion in Corollary A.4 and Remark A.5 yield,

$$\begin{aligned} h^{rB} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_0^* t_n \tilde{\beta}_{n,t}^{(1)} &= h^{rB} \sqrt{n} \mathbf{P}_n \zeta \varepsilon \tilde{\beta}_{n,t}^{(1)} + \\ h^{rB} \sqrt{n} \mathbf{P}_n \zeta (r - r^{[B]}) t_n \tilde{\beta}_{n,t} &+ h^{rB} \sqrt{n} \mathbf{P}_n (r^{[B]} - r^{*[0]}) t_n \tilde{\beta}_{n,t} + o_P^*(1). \end{aligned}$$

The first term on the RHS of the previous display is  $o_P^*(1)$  uniformly over  $\mathbb{T}$ . Using the rates in Lemma C.6, and Assumption 7, we obtain  $\|(r^{*[0]} - r)t_n\|_\infty = o_P^*(n^{-1/4})$  and  $\|(r^{[B]} - r)t_n\|_\infty = o_P(n^{-1/4})$ . By these rates and Assumption 5, the second and the third term on the RHS of the above display are  $o_P^*(1)$  uniformly over  $\mathbb{T}$ , so we obtain the result of the present Lemma.  $\square$

Let us recall that  $F_0$  denotes the cdf generated by the random variable  $\int \cdot \mathbf{G}^2 \mu(dt)$  defined in Corollary A.4.

**Corollary B.4.** (Proof of Proposition 6.1) Under Assumption 1-6, uniformly over  $\mathbb{T}$ ,

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* \hat{t}_n \varphi_t = \sqrt{n} \mathbf{P}_n \zeta \varepsilon \varphi_t^\perp + o_P^*(1).$$

Accordingly, for every continuity point of  $F_0$ ,

$$\mathbf{P}^\zeta(S_n^* \leq z) \xrightarrow{P} F_0(z) \text{ and } \mathbf{P}^\zeta(S_n^* \leq z) - F_0(z) \xrightarrow{P} 0.$$

*Proof.* The first result of the present Corollary follows from Lemma B.2 and B.3. For the second result, define  $\tilde{S}_n^* = \int \sqrt{n} \mathbf{P}_n \zeta \varepsilon \varphi_t^\perp \cdot \mathbf{G}^2 \mu(dt)$ . From the expansion of the present Corollary and the continuity of the functional  $\int |\cdot|^2 \mu(dt)$ ,

$$\tilde{S}_n^* - S_n^* = o_P^*(1). \quad (21)$$

Theorem 3.6.13 in van der Vaart & Wellner (1996) ensures that

$$\sqrt{n} \mathbf{P}_n \zeta \varepsilon \varphi_t^\perp \cdot \mathbf{G} \text{ for almost all trajectories.}$$

Using again the continuity of the functional  $\int |\cdot|^2 \mu(dt)$  and the weak convergence in the previous display, by an application of the Continuous Mapping Theorem we obtain  $\tilde{S}_n^* \xrightarrow{P} \int \cdot \mathbf{G}^2 \mu(dt)$ .

Hence,

$$\mathbb{P}^\zeta(\tilde{S}_n^* \leq z) \xrightarrow{P} F_0(z) \text{ for all continuity points of } F_0.$$

Now, let  $z$  be a continuity point of  $F_0$ , and fix an arbitrary  $\delta > 0$  such that also  $z - \delta$  and  $z + \delta$  are continuity points of  $F_0$ . Then, by Eq. 21,

$$\begin{aligned} \mathbb{P}^\zeta(S_n^* \leq z) &= \mathbb{P}^\zeta(\tilde{S}_n^* \leq z + (\tilde{S}_n^* - S_n^*) \cap S_n^* - \tilde{S}_n^* \leq \delta) + \mathbb{P}^\zeta(S_n^* - \tilde{S}_n^* > \delta) \\ &\leq \mathbb{P}^\zeta(\tilde{S}_n^* \leq z + \delta) + o_P(1) \end{aligned}$$

A similar reasoning yields,

$$\mathbb{P}^\zeta(\tilde{S}_n^* \leq z - \delta) + o_P(1) \leq \mathbb{P}^\zeta(S_n^* \leq z).$$

Putting together the previous three displays, since  $z$ ,  $z - \delta$ , and  $z + \delta$  are all continuity points of  $F_0$ ,

$$F_0(z - \delta) + o_P(1) \leq \mathbb{P}^\zeta(S_n^* \leq z) \leq F_0(z + \delta) + o_P(1)$$

Since  $z$  is a continuity point of  $F_0$ , we can choose a  $\delta$  small enough such that  $|F_0(z - \delta) - F_0(z)| \leq s$  and  $|F_0(z + \delta) - F_0(z)| \leq s$ . Hence, the above display and such definition of  $\delta$ ,

$$-s + o_P(1) \leq \mathbb{P}^\zeta(S_n^* \leq z) - F_0(z) \leq s + o_P(1).$$

The above  $s$  can be chosen to be arbitrarily small, so  $\mathbb{P}(\tilde{S}_n^* \leq z) \xrightarrow{P} F_0(z)$ . From Proposition 5.1  $\hat{F}_n(z) \xrightarrow{P} F_0(z)$  for every continuity point of  $F$ , hence we conclude.  $\square$

## C Auxiliary Lemmas

**Lemma C.1. (Entropy bounds)** Let  $\Psi := \{x \mapsto \psi_t(x) : t \in \mathbb{T}\}$  be a class of functions with the mapping  $(t, x) \mapsto \psi_t(x)$  satisfying the Lipschitz condition  $\|\psi_{t_1} - \psi_{t_2}\|_\infty \leq C\|t_1 - t_2\|$  for all  $t_1, t_2 \in \mathbb{T}$ . For  $\mathbf{C}_n := \mathbf{C}(\mathbf{W}_n)$ , with  $\mathbf{W}_n$  and  $\mathbf{C}(\mathbf{W})$  defined in Eq. 5 and 6, it holds that:

$$(i) N(\delta, \Psi \mathbf{C}_n t_n, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{C}_n, \|\cdot\|_\infty, \mathbf{W}_n) \cdot N(C\delta, \mathbb{T}, \|\cdot\|)$$

$$(ii) \log N(\delta, \mathbf{C}_n t_n, \|\cdot\|_\infty) \leq C\delta^{-v} \text{ with } v \in (0, 2).$$

*Proof.* By definition of  $\mathbf{X}_n$  and  $\mathbf{C}_n$  and since  $t_n(w) = \mathbf{1}\{f(w) \geq \tau_n\}$ , it is readily obtained

that  $N(\delta, \mathbf{C}_n t_n, \|\cdot\|_\infty) \leq N(\delta, \mathbf{C}_n, \|\cdot\|_{\infty, X_n})$ . Also, given two classes  $\mathbf{F}$  and  $\mathbf{A}$  of uniformly bounded functions defined on an arbitrary set  $\mathbf{Y}$ , it is immediate to see that  $N(\delta, \mathbf{F} \mathbf{A}, \|\cdot\|_{\infty, \mathbf{Y}}) \leq N(C\delta, \mathbf{F}, \|\cdot\|_{\infty, \mathbf{Y}}) \cdot N(C\delta, \mathbf{A}, \|\cdot\|_{\infty, \mathbf{Y}})$  for a finite constant  $C$ . Using the Lipschitz property of the class  $\Psi$  and by proceeding as in Example 19.7 of [van der Vaart \(1998\)](#),  $N(\delta, \Psi, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{T}, \|\cdot\|)$ . The previous three inequalities imply that  $N(\delta, \Psi \mathbf{C}_n t_n, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{C}_n, \|\cdot\|_{\infty, W_n}) \cdot N(\delta, \mathbf{T}, \|\cdot\|)$ . Point (ii) therefore follows from this last inequality, the compactness of  $\mathbf{T}$  and Theorem 2.7.1 in [van der Vaart & Wellner \(1996\)](#) (see also Observation in Section 5).  $\square$

**Lemma C.2. (Trimming)** Under Assumption 4 and 6(i),

$$(i) \sqrt{n} \mathbf{P}_n |\hat{t}_n - t_n| = o_P(1)$$

(ii) If  $\sup_{t \in \mathbf{T}} \|\hat{g}_t \hat{t}_n\|_\infty = o_P(1)$  and  $\sup_{t \in \mathbf{T}} \|\hat{g}_t t_n\|_\infty = o_P(1)$ , then  $\sqrt{n} \mathbf{P}_n \hat{g}_t (\hat{t}_n - t_n) = o_P(1)$  uniformly over  $\mathbf{T}$ .

*Proof.* Let  $\mathbf{B}_n := \{ \|\hat{f}^{[0]} - f\|_\infty \leq C d_{n,f} \}$  and fix  $\delta > 0$  arbitrarily small. Assumption 6(i) ensures that, by choosing  $C$  large enough,  $P(\mathbf{B}_n) > 1 - \delta$  for any large  $n$ . By definition of  $\hat{t}_n$  we can write

$$\hat{t}_n(w) = 1' f(w) \geq \tau_n \left( 1 - \frac{\hat{f}^{[0]}(w) - f(w)}{\tau_n} \right).$$

If the event  $\mathbf{B}_n$  holds and  $n$  is large enough so that  $C d_{n,f} / \tau_n \leq 1/2$ ,

$$1 - C \frac{d_{n,f}}{\tau_n} \leq 1 - \frac{\hat{f}^{[0]}(w) - f(w)}{\tau_n} \leq 1 + C \frac{d_{n,f}}{\tau_n} \leq \frac{3}{2}$$

for all  $x \in \mathbf{X}$ . For such  $n$  and when the event  $\mathbf{B}_n$  holds, by the two previous displays

$$f(w) \geq \frac{3}{2} \tau_n \Rightarrow \hat{t}_n(w) = t_n(w) = 1.$$

Hence, when  $\mathbf{B}_n$  holds and  $C d_{n,f} / \tau_n \leq 1/2$ , we have  $|\hat{t}_n - t_n|(w) \leq 1' f(w) \leq \frac{3}{2} \tau_n$ . Using this, Markov's inequality, and Assumption 4,

$$\begin{aligned} P(\sqrt{n} \mathbf{P}_n |\hat{t}_n - t_n| > \delta) &\leq P(\sqrt{n} \mathbf{P}_n |\hat{t}_n - t_n| > \delta \cap \mathbf{B}_n) + \delta \leq \\ &\delta^{-1} \sqrt{n} \cdot P(f(W) \leq \frac{3}{2} \tau_n) + \delta = o(1) + \delta. \end{aligned}$$

By the arbitrariness of  $\delta$  we conclude for point (i). Point (ii) follows immediately from point (i) and after noticing that  $\hat{t}_n - t_n = \frac{2}{n} (t + \hat{t}_n)(t_n - \hat{t}_n)$ .  $\square$

**Lemma C.3.** Assume that  $Z, P_Z, F_n$ , and  $\Psi$  are as in Lemma C.1. Let  $f_0$  be a fixed function defined over  $Z$ , and  $\hat{f}$  be a random function over  $Z$ , where the randomness is considered wrt the probability  $P_Z$ . Define  $G_n := \sqrt{n}(P_n - P_Z)$ . If

$$(i) \|\hat{f} - f_0\|_{L_2(P_Z)} = o_P(1),$$

$$(ii) P(\hat{f} \in F_n) \rightarrow 1, \text{ with } \log N_{[\cdot]}(\delta, F_n, L_2(P_Z)) \leq C\delta^{-v} \text{ and } v \in (0, 2),$$

then

$$G_n(\hat{f} - f_0)\psi_t = o_P(1) \text{ uniformly over } T.$$

*Proof.* Define  $\tilde{F}_n := (F_n - f_0)\Psi$ . Since the entropy of  $F_n - f_0$  is equal to that of  $F_n$ , Lemma C.1 implies that  $\log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$ . Fix  $\delta > 0$ . By Assumptions (i) and (ii) of the present Lemma, for an arbitrary  $\eta > 0$ ,

$$\limsup_{n \rightarrow \infty} P \left( \sup_{t \in T} G_n(\hat{f} - f_0)\psi_t > \eta \right) \leq \limsup_{n \rightarrow \infty} P \left( \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta \right),$$

where  $\tilde{F}_n(\delta) := \{f \in F_n; \|f - f_0\|_{L_2(P_Z)} < \delta\}$ . The RHS of the previous display can be upperbounded according to the maximal inequality in Lemma 19.34 of [van der Vaart \(1998\)](#). Since  $\log N_{[\cdot]}(\delta, \tilde{F}_n(\delta), L_2(P_Z)) \leq \log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$  with  $v \in (0, 2)$ , we can choose a small enough  $\delta$  to make such upperbound arbitrarily small. By the arbitrariness of  $\eta$ , we conclude.  $\square$

**Lemma C.4.** Assume that  $Z, P_Z, F_n, \Psi$ , and  $f_0$  are as in Lemma C.3. Let  $\hat{f}$  be a random function over  $Z$  where the randomness is considered wrt a probability  $P = P_Z \otimes P_\xi^*$ , with  $P_\xi^*$  being a probability measure. Define  $G_n := \sqrt{n}(P_n - P)$ . If

$$(i) \|\hat{f} - f_0\|_{L_2(P)} = o_{P_\xi^*}(1),$$

$$(ii) P_U^*(\hat{f} \in F_n) \xrightarrow{P} 1, \text{ with } \log N_{[\cdot]}(\delta, F_n, L_2(P_Z)) \leq C\delta^{-v} \text{ and } v \in (0, 2),$$

then

$$G_n(\hat{f} - f_0)\psi = o_{P_U^*}(1) \text{ uniformly over } T.$$

*Proof.* By the same reasoning as in Lemma C.3,  $\log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$ . Fix  $\delta > 0$ . Assumptions (i) and (ii) of the present Lemma ensure that for an arbitrary  $\eta > 0$ ,

$$\begin{aligned} & P_{\xi^*} \left( \sup_{t \in T} G_n(\hat{f} - f_0)\psi_t > \eta \right) \leq \\ & P_{\xi^*} \left( \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta \right) + P_{\xi^*} \left( \sup_{t \in T} \|(\hat{f} - f_0)\psi_t\|_{L_2(P)} > \delta \right) + P_{\xi^*} \left( \hat{f} \notin F_n \right) = \\ & P_{\xi^*} \left( \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta \right) + o_P(1). \end{aligned}$$

For the first term on the RHS of the above display, using Markov's inequality

$$P_Z \cdot P_\xi^* \sup_{f \in \tilde{\mathcal{F}}_n(\delta)} G_n f \cdot \eta \leq \frac{1}{\eta} E_{P_Z} \cdot P_\xi^* \sup_{f \in \tilde{\mathcal{F}}_n(\delta)} G_n f \cdot \eta = \frac{1}{\eta} P \sup_{f \in \tilde{\mathcal{F}}_n(\delta)} G_n f \cdot \eta,$$

where the last inequality follows from the product structure of  $P$ , i.e.  $P = P_Z \otimes P_\xi^*$ , and Tonelli-Fubini's Theorem. With the same arguments as in the proof of Lemma C.3, we can show that by choosing  $\delta$  small enough the RHS of the previous display can be made arbitrarily small. So, by the arbitrariness of  $\eta$  and the previous two displays we conclude.  $\square$

**Lemma C.5.** Let  $\{Z_{i,n}\}_{i=1}^n$  be a triangular array of real-valued random variables such that  $|Z_{n,i}| \leq C$  for all  $n$  and  $i$ , and let  $L$  be a kernel function that is Lipschitz continuous. If  $\frac{\log n}{nh^d} = o(1)$ , then

$$\therefore \frac{1}{nh^d} \sum_{i=1}^n Z_{n,i} L \left( \frac{W_i - w}{h} \right) = h^{-d} E \left[ Z_{n,i} L \left( \frac{W_i - w}{h} \right) \right] + o_p \left( \frac{\log n}{nh^d} \right).$$

*Proof.* The proof is the same as Theorem 1.4 in [Li & Racine \(2006\)](#) (pages 36-40).  $\square$

**Lemma C.6.** Under Assumption 5 and 6(i),

$$(i) \|\hat{T}_{\hat{\varepsilon}_b}\|_{\infty, X} = O_p \left( \frac{1}{\tau_n^b} \frac{\log n}{nh^d} + \frac{\|(r^{[0]} - r)\hat{t}_n\|_{\infty, X}}{\tau_n^b} + \frac{P_n \|\hat{t}_n - t\|_n}{h^d \tau_n^b} \right) \\ (ii) \|(r^{[b]} - r)\hat{t}_n\|_{\infty} = O_p \left( \frac{1}{\tau_n^b} \frac{\log n}{nh^d} + \frac{\|(r^{[0]} - r)\hat{t}_n\|_{\infty, X}}{\tau_n^b} + \frac{P_n \|\hat{t}_n - t\|_n}{h^d \tau_n^b} \right).$$

The same rates hold for  $\|(r^{[b]} - r)\hat{t}_n\|_{\infty}$ .

*Proof.* By definition of  $\hat{T}_{\hat{\varepsilon}_b}$  and by using Lemma C.5,

$$\begin{aligned} \left\| \frac{F(w)}{\varepsilon_0} \right\|_{\infty} &\leq \left\| \frac{1}{nh^d} \sum_{i=1}^n \varepsilon_i \hat{t}_{n,i} K \left( \frac{W_i - w}{h} \right) \right\|_{\infty} + \left\| \frac{1}{nh^d} \sum_{i=1}^n (r_i - \hat{r}_i^{[0]}) \hat{t}_{n,i} K \left( \frac{W_i - w}{h} \right) \right\|_{\infty} \leq \\ &\leq \left\| \frac{1}{nh^d} \sum_{i=1}^n \varepsilon_i \hat{t}_{n,i} K \left( \frac{W_i - w}{h} \right) \right\|_{\infty} + \left\| \frac{1}{nh^d} \sum_{i=1}^n \varepsilon_i (\hat{t}_{n,i} - t_{n,i}) K \left( \frac{W_i - w}{h} \right) \right\|_{\infty} + O_p \left( \|(r^{[0]} - r)\hat{t}_n\|_{\infty} \right) = \\ &= O_p \left( \frac{\log n}{nh^d} \right) + O_p \left( \frac{P_n \|\hat{t}_n - t\|_n}{h^d} \right) + O_p \left( \|(r^{[0]} - r)\hat{t}_n\|_{\infty} \right) \quad (22) \end{aligned}$$

uniformly over  $W$ . By Eq. 22 and using the recursive structure of  $r^{[d]}$  in Eq. 4,

$$\|(r^{[1]} - r)\hat{t}_n\|_\infty = \|(r^{[0]} - r)\hat{t}_n + \frac{\hat{T}_{\hat{\varepsilon}_0} \cdot}{\hat{f}^{[0]} \cdot \cdot \infty} = O_P \cdot \frac{\|(r^{[0]} - r)\hat{t}_n\|_\infty}{\tau_n} + \frac{1}{\tau_n} \cdot \frac{\overline{\log n}}{nh^d} + \frac{P_n |\hat{t}_n - t_n|}{h^d \tau_n} \Sigma \quad (23)$$

The above display and a reasoning similar to Eq. 22 yields

$$\|\hat{T}_{\hat{\varepsilon}}\|_\infty = O_P \cdot \frac{\|(r^{[0]} - r)\hat{t}_n\|_\infty}{\tau_n} + \frac{1}{\tau_n} \cdot \frac{\overline{\log n}}{nh^d} + \frac{P_n |\hat{t}_n - t_n|}{h^d \tau_n} \Sigma \quad (24)$$

We can now show the result of the present lemma by a simple induction. By proceeding as in Eq. 24 and 23, if for  $b < B$

$$\|(r^{[b]} - r)\hat{t}_n\|_\infty = O_P(l_n^{[b]}) \text{ and } \|\hat{T}_{\hat{\varepsilon}_b}\|_\infty = O_P(l_n^{[b]}) \quad (25)$$

$$\text{for } l_n^{[b]} = \frac{\|(r^{[0]} - r)\hat{t}_n\|_\infty}{\tau_n^b} + \frac{1}{\tau_n^b} \cdot \frac{\overline{\log n}}{nh^d} + \frac{P_n |\hat{t}_n - t_n|}{h^d \tau_n^b}$$

then the same property will also hold for  $b+1$ . Since Eq. 23 and 24 ensure that 25 holds for  $b=1$ , we conclude.  $\square$

**Lemma C.7. (Convergence Rates for the Bootstrap)** Under Assumptions 5 and 6(i),

$$(i) \quad F_{\hat{\varepsilon}_b}^* \cdot \cdot \infty = O_P \cdot \tau_n^{-b} \cdot \frac{\overline{\log n}}{nh^d} + \hat{l}_n^{[b]} + \tau_n^{-b} \|(r^{*[0]} - r^{[0]})\hat{t}_n\|_\infty \text{ for } b = 0, 1, \dots, B-1;$$

$$(ii) \quad \|(r^{*[b]} - r^{[0]})\hat{t}_n\|_\infty = O_P \cdot \tau_n^{-b} \cdot \frac{\overline{\log n}}{nh^d} + \hat{l}_n^{[b]} + \tau_n^{-b} \|(r^{*[0]} - r)\hat{t}_n\|_\infty \text{ for } b = 1, \dots, B;$$

where  $\hat{l}_n^{[b]} = \frac{P_n |\hat{t}_n - t_n|}{\tau_n^b h^d} + \frac{\|(r^{[B]} - r)\hat{t}_n\|_\infty}{\tau_n^b} + \tau_n^{-b} \cdot \frac{\hat{T}_{\hat{\varepsilon}_B} \cdot}{f} \hat{t}_n \cdot \cdot \infty$ . The same rates also hold for  $\|(r^{*[b]} - r^{[0]})\hat{t}_n\|_\infty$ .

*Proof.* The proof is similar to the proof of Lemma C.6. We start with  $b=0$ . By using Lemma C.5,

$$\begin{aligned} F_{\hat{\varepsilon}_0}^*(w) &= \frac{1}{nh^d} \sum_{i=1}^n \xi_i (Y_i - \hat{r}_i^{[B]}) \hat{t}_{n,i} K \cdot \frac{W_{i-w}}{h} \Sigma + \frac{1}{nh^d} \sum_{i=1}^n (\hat{r}_i^{[B]} - \hat{r}_i^{*[0]}) \hat{t}_{n,i} K \cdot \frac{W_{i-w}}{h} \Sigma \leq \\ &= \frac{1}{nh^d} \sum_{i=1}^n \xi_i \varepsilon_{t,i} K \cdot \frac{W_{i-w}}{h} \Sigma + \frac{1}{nh^d} \sum_{i=1}^n \xi_i \varepsilon_i (\hat{t}_{n,i} - t_{n,i}) K \cdot \frac{W_{i-w}}{h} \Sigma + \\ &\quad + \frac{1}{nh^d} \sum_{i=1}^n \xi_i (r_i - \hat{r}_i^{[B]}) \hat{t}_{n,i} K \cdot \frac{W_{i-w}}{h} \Sigma + \\ &+ \frac{1}{nh^d} \sum_{i=1}^n (\hat{r}_i^{[0]} - \hat{r}_i^{*[0]}) \hat{t}_{n,i} K \cdot \frac{W_{i-w}}{h} \Sigma + \frac{1}{nh^d} \sum_{i=1}^n \sum_{q=1}^{B-1} \frac{\hat{T}_{\hat{\varepsilon}_q}(W_i)}{f} \cdot \hat{t}_{n,i} K \cdot \frac{W_{i-w}}{h} \Sigma = \\ &= O_P \cdot \frac{\overline{\log n}}{nh^d} + O_P \cdot \frac{P_n |\hat{t}_n - t_n|}{h^d} + O_P \cdot \|(r - \hat{r}^{[B]})\hat{t}_n\|_\infty + \\ &\quad + O_P \cdot \|(r^{[0]} - r^{*[0]})\hat{t}_n\|_\infty + O_P \cdot \frac{\hat{T}_{\hat{\varepsilon}_B} \cdot}{f} \hat{t}_n \cdot \cdot \infty, \end{aligned} \quad (26)$$



uniformly over  $W$ . From the previous display and since  $\hat{r}^{*[1]} = \hat{r}^{*[0]} + \frac{\hat{T}_{\varepsilon_0}^*}{f}$ ,

$$\begin{aligned} ||(\hat{r}^{*[1]} - \hat{r}^{*[0]})\hat{t}_n||_\infty &\leq ..\hat{T}_{\varepsilon_0}^*.._\infty + ||(\hat{r}^{*[0]} - \hat{r}^{[0]})\hat{t}_n||_\infty = \\ O_P \left( \tau_n^{-1} \cdot \overline{\frac{\log n}{nh^d}} + \hat{l}_n^{[1]} + \tau_n^{-1} ||(\hat{r}^{[0]} - \hat{r}^{*[0]})\hat{t}_n||_\infty \right) &\quad \Sigma \end{aligned} \quad (27)$$

The above display and a reasoning similar to Eq. 26 yield

$$..\hat{T}_{\varepsilon_1}^*.._\infty = O_P \left( \tau_n^{-1} \cdot \overline{\frac{\log n}{nh^d}} + \hat{l}_n^{[1]} + \tau_n^{-1} ||(\hat{r}^{[0]} - \hat{r}^{*[0]})\hat{t}_n||_\infty \right) \quad \Sigma \quad (28)$$

The result of the Lemma now follows from a simple induction. In particular, assume that for  $b < B$ :

$$\begin{aligned} ..\hat{T}_{\varepsilon_b}^*.._\infty &= O_P \left( \tau_n^{-b} \cdot \overline{\frac{\log n}{nh^d}} + \hat{l}_n^{[b]} + \tau_n^{-b} ||(\hat{r}^{*[0]} - \hat{r}^{[0]})\hat{t}_n||_\infty \right) \quad \Sigma \\ ||(\hat{r}^{*[b]} - \hat{r}^{[0]})\hat{t}_n||_\infty &= O_P \left( \tau_n^{-b} \cdot \overline{\frac{\log n}{nh^d}} + \hat{l}_n^{[b]} + \tau_n^{-b} ||(\hat{r}^{*[0]} - \hat{r}^{[0]})\hat{t}_n||_\infty \right) \quad \Sigma \end{aligned} \quad (29)$$

Then, proceeding in the same way as in Eq. 26 and 27 shows that the property will hold also for  $b+1$ . By Eq. 27 and 28 the Induction Assumption in 29 holds for  $b=1$ , so we conclude.  $\square$

# Chapter 3: Testing Bayesian-Nash Behavior in Binary Games with Incomplete Information and Correlated Types

Elia Lapenta and Pascal Lavergne\*

## Abstract

We provide a test to check if the distribution of the observed data can be rationalized by a unique Bayesian-Nash equilibrium of a binary game with incomplete information, where agents' types can be mutually correlated. Testing this assumption is useful for two reasons. First, the uniqueness of the Bayesian-Nash equilibrium is key to identify the fundamentals of the game. Second, testing for a Bayesian-Nash behavior is interesting *per se*, as it is an assumption often postulated in game-theoretical models. The test we propose relies on rationalization results in [Liu et al. \(2017\)](#). We construct our test statistic by an  $L_2$ -boosting procedure. This is effective to control the estimation bias arising in our context. Since the asymptotic distribution of our statistic depends on unknown features of the data, we propose a novel Multinomial Bootstrap procedure to obtain the critical value and prove its validity. This procedure resamples the observations by imposing that a unique Bayesian-Nash equilibrium is played. A Monte Carlo experiment shows the good small-sample performance of the test.

**Keywords:** Hypothesis Testing, Empirical Games, Bayesian-Nash Equilibrium, Bootstrap, Boosting.

**JEL Classification:** C01, C10, C12, C14, C57

## 1 Introduction

Game-theoretical models are often used in economics to describe strategic interactions between different agents. Games of complete information, for instance, have been employed to model labor force participation ([Bjorn & Vuong, 1985](#); [Kooreman, 1994](#); [Soetevent & Kooreman, 2007](#)), or firms' entry decisions in oligopolistic frameworks ([Bresnahan & Reiss, 1991](#); [Bresnahan & Reiss, 1990](#); [Berry, 1992](#); [Ciliberto & Tamer, 2009](#); [Jia, 2008](#)). These setups assume that each player observes all the features of the other players' profit functions. Differently, games of incomplete information relax this hypothesis by allowing each player to have a private information that the other agents cannot observe. Such frameworks have been employed to describe firms' radio commercial decisions ([Sweeting, 2009](#)), capital

---

\*Toulouse School of Economics

investment strategies ([Bajari et al., 2010a](#); [Aradillas-Lopez, 2010](#)), the decision of opening grocery stores ([Seim, 2006](#)), and social interactions ([Brock & Durlauf, 2007](#)). In the case of games with incomplete information, the efforts of the literature have mostly focused on identification and estimation ([Brock & Durlauf, 2007](#); [Aguirregabiria & Mira, 2007](#); [Sweeting, 2009](#); [Tang, 2010](#); [Lewbel & Tang, 2015a](#)). In these contexts it is key to impose that a unique Bayesian-Nash equilibrium is played ([Brock & Durlauf, 2007](#); [Aradillas-Lopez, 2012](#); [Bajari et al., 2010a](#); [Lewbel & Tang, 2015b](#); [Liu et al., 2017](#)). However, little has been done to check the validity of such an assumption. The only paper that attempts at testing whether the distribution of the observed data is coherent with a unique Bayesian-Nash equilibrium is [de Paula & Tang \(2010\)](#). Since the authors provide a test based on a fixed value of the explanatory variable, their results are conditional on such values. Furthermore, they rely on the independence of agents' types.

This paper contributes to the literature on incomplete information games by proposing a novel test to check if the distribution of the observed data can be rationalized (or characterized) by a unique Bayesian-Nash equilibrium of a game with incomplete information. In this game agents play binary strategies and their private types are allowed to be mutually correlated. Checking such hypothesis can be interesting from different perspectives. First, it can serve as a test to check the validity of the Bayesian-Nash assumption. Second, it can be employed to check if the uniqueness of the equilibrium is coherent with the observed data. To build the test we rely on results in [Liu et al. \(2017\)](#). The authors derive a characterization of the Bayesian-Nash hypothesis in terms of restrictions on the distribution of the observed data. Such a result does not rely neither on the mutual independence of agents' private information nor on the functional forms of agents' payoffs. Hence, the test we propose is fully nonparametric. This feature is attractive, as in the presence of parametric restrictions on either agents' payoff or their private information, the conclusions reached by any test would be conditional on the functional forms imposed. Moreover, since the result by [Liu et al. \(2017\)](#) does not rely on the mutual independence of agents' private information, the test is also robust to the presence of correlated private types. Allowing for such correlation might be relevant in empirical contexts. For instance, in oligopolistic entry games agents' private information might be mutually correlated because of unobserved firms' profitability, see [Berry \(1992\)](#), or in network formation games the correlation among agents' private types might arise because of homophily, see [Mcpherson et al. \(2001\)](#). Finally, we do not condition our test on fixed values of the explanatory variables.

The characterization of [Liu et al. \(2017\)](#) we start from involves a nonparametric transformation of the conditional probabilities concerning agents' decisions. The conditioning variables can be interpreted as the exogenous covariates of the game. We propose a test statistic based on a two-step procedure. In a first step, we estimate the conditional probabilities concerning the decisions of each agent. In a second step, we estimate the nonparametric transformation based on these conditional probabilities. To handle the bias arising from this two-step nonparametric estimation, we use an  $L_2$  boosting procedure. This can be interpreted as an iterative bias-correction method. Such an algorithm has been originally proposed in the machine-learning literature and has been shown to be quite effective in handling the bias arising from the nonparametric estimation, see [Di Marzio & Taylor \(2008\)](#), [Park et al. \(2009\)](#), [Cornillon et al.](#)

(2014). To the best of our knowledge, the implementation of such an algorithm in a testing problem involving multi-step estimation is novel in the literature. The statistic we construct is expressed as a weighted sum of the residuals and has simple closed-form expression. We show that under the null hypothesis (i.e. that the distribution of the data is rationalized by a unique Bayesian-Nash equilibrium) it converges to a transformation of a Gaussian process.

Since the asymptotic distribution depends on unknown features of the data, we propose to compute the critical values by a new Multinomial Bootstrap procedure. This incorporates all the restrictions the null hypothesis imposes on the distribution of the data, assuming that a unique Bayesian-Nash equilibrium is played. For the practical implementation of the test we use kernel methods, so that both the first step and the second step estimations are carried on by locally-constant regressions. The advantage of the  $L_2$ -boosting procedure in this case is to allow the implementation of the test without undersmoothing. Moreover, it considerably enlarges the set of bandwidths and kernels admissible for the test.

To the best of our knowledge, this is the first paper aiming at testing the rationalization of the data by a unique Bayesian-Nash equilibrium with possibly correlated agents.

**Related literature.** Beyond the literature on incomplete information games, this paper is also related to the literature on estimation and testing in the presence of generated regressors. Pagan (1984) addresses the problem of estimation in the presence of generated variables. Ahn & Powell (1993), Newey (1994), Ahn (1997), and Newey *et al.* (1999), and are other early contributions to estimating semiparametric models with generated covariates. Mammen *et al.* (2012), Mammen *et al.* (2016), Blundell & Powell (2004), Rothe (2009), Vanhems & Keilegom (2019), Escanciano *et al.* (2016), Escanciano *et al.* (2014), Hahn *et al.* (2018), Hahn & Ridder (2013) are finally more recent works analyzing estimation in setups with generated variables. However, the problem of testing a hypothesis like the one at the center of this paper is not addressed.

Our paper is also related to the extensive literature on specification testing, see Fan & Li (1996), Lavergne *et al.* (2015), Lavergne & Vuong (1996), Delgado & Manteiga (2001), Stinchcombe & White (1998), among others. We contribute to this literature in three ways. First, these papers assume that all the variables are observed, while we have generated variables. Second, we use an  $L_2$  boosting procedure that gives rise to a recursive structure in our statistic. Third, we propose a new Multinomial Bootstrap scheme for the computation of the critical values.

**Organization of the paper.** The reminder of the paper goes as follows. In Section 2, we describe the game-theoretical framework and the rationalization result of Liu *et al.* (2017). Section 3 builds the test, and the following Section 4 describes the application of the  $L_2$  boosting algorithm to the case at hand. The assumptions and the asymptotic behavior of the test statistic are reported in Section 5. Section 6 presents the Multinomial Bootstrap procedure, describes its implementation, and shows its validity. Section 7 assesses the small-sample properties of our test in a Monte Carlo simulation study. Finally,

Section 8 concludes. All technical proofs are reported in the Appendix.

## 2 Rationalization of Incomplete Information Games with Correlated Agents

**Basic features of the model.** We assume the presence of  $S$  players and denote by  $s$  the generic player, so that  $s = 1, \dots, S$ . Each agent can take a binary action  $a_s \in \{0, 1\}$ . Let  $a = (a_1, \dots, a_S)$  be an action profile,  $\mathbf{A} = \{0, 1\}^S$  be the space of action profiles,  $a_{-s} = (a_1, \dots, a_{s-1}, a_{s+1}, \dots, a_S)$  be the vector containing the actions of all the players but  $s$ , and  $\mathbf{A}_{-s} = \{0, 1\}^{S-1}$  be the set containing all the possible values of  $a_{-s}$ . We denote with  $x$  a specific value of the vector of the exogenous covariates observed by each player and the econometrician. Each agent has a private information -or type- that is unknown to both the other players and the researcher. Let  $\theta_s$  stand for the agent  $s$ ' type or private information. The players other than  $s$  do not know  $\theta_s$ , but they know the distribution of  $\theta = (\theta_1, \dots, \theta_S)$  conditional on  $X$ , denoted by  $F_{\theta|X}$ . Agents' payoff from taking action 1 is  $\pi_s(a_{-s}, X) - \theta_s$ , while we normalize to zero the payoff from taking action 0. Hence, the pay-off function of agent  $s$  can be written as

$$\Pi_s(a_s, a_{-s}, X, \theta_s) = a_s x \{ \pi_s(a_{-s}, X) - \theta_s \}.$$

To fix the ideas at this stage, the game just described can be thought of as a free-entry game between different firms. Each firm owns a private information unknown to the other competitors and must decide whether to enter or not to enter a specific market.  $a_s = 1$  denotes the entry in a specific market by firm  $s$ , while  $a_s = 0$  denotes that the firm  $s$  has decided not to enter the market. This is a basic example and later on we will provide other applications.

We now impose the following assumptions:

**Assumption E (Exogeneity).**  $\theta = (\theta_1, \dots, \theta_S)$  is independent from  $X$ .

**Assumption D (Density).**  $\theta = (\theta_1, \dots, \theta_S)$  admits a density  $f_{\theta}$  with respect to the Lebesgue measure that is continuously differentiable.

The latter assumption is just a regularity condition. The first assumption is usually made in the empirical literature on discrete games and has been often used for the identification of the structural part of the profit functions, see [Bajari et al. \(2010b\)](#). Notice that it is not imposing the mutual independence of agents' types. This is an important feature of the framework. The presence of correlation between players' types can capture heterogeneous effects, the presence of homophily between several agents, or the part of profitability not explained by the interaction between firms. As regards the first, consider the free-entry game introduced above, where several firms must decide whether to enter or not to enter a specific market. Assume that the econometrician has a data set consisting of a cross-section of

markets/industries. The presence of correlation between agents' types, i.e. between those components of the model which are not observed by the econometrician, can handle a random effect due to the unobserved heterogeneity between several markets/industries. As regards homophily, the presence of correlated types seems to be a reasonable assumptions when we model peers' decisions about friendship. It is reasonable to think that each agent would be more likely to establish a friendship relationship with those who are more similar to him, with these similarities being explained by those components of the model unobserved to the econometrician.

**Equilibrium.** The equilibrium we consider in this paper is of the Bayesian-Nash type. The equilibrium pure-strategy for agent  $s$  can be seen as a function of the covariates  $X$  and the type  $\theta_s$ . So, for a certain collection of mappings  $\{\delta_s, s = 1, \dots, S\}$  the equilibrium action of each player will take the form

$$a_s = \delta_s(X, \theta_s).$$

The collection of functions  $\delta = (\delta_1, \dots, \delta_S)$  stands for a profile of strategies. For any given profile  $\delta$ , let  $\sigma^\delta(a_{-s}, X)$  be the probability that the players other than  $s$  take the actions  $a_{-s} \in \mathbf{A}_{-s}$ . Using Assumptions [E](#) and [D](#), we can write

$$\sigma^\delta(a_{-s}|X, \theta_s) = \int \mathbf{1}\{\delta_q(X, \theta_q) = a_q \text{ for } q \neq s\} \cdot f_\theta(\theta_1, \dots, \theta_S|\theta_s) d\theta_1, \dots, \theta_{s-1}, \theta_{s+1}, \dots, \theta_S.$$

Assuming that player  $s$  believes the other agents will play according to the strategies contained in  $\delta$ , player  $s$  will play action 1 if and only if her expected profit is larger than zero. So player  $s$ ' actions in equilibrium can be described by the following equivalence:

$$a_s = 1 \text{ if and only if } \sum_{a_{-s} \in \mathbf{A}_{-s}} \pi_s(a_{-s}, X) \cdot \sigma_\delta(a_{-s}|X, \theta_s) - \theta_s \geq 0.$$

This implies that a profile of strategies  $\delta$  will be a Bayesian-Nash equilibrium for the game just described if and only if it solves the system

$$\delta_s(X, \theta_s) = 1, \quad \sum_{a_{-s} \in \mathbf{A}_{-s}} \pi_s(a_{-s}, X) \cdot \sigma_\delta(a_{-s}|X, \theta_s) - \theta_s \geq 0, \quad \forall s = 1, \dots, S \text{ and for all } X, \theta_s. \quad (1)$$

The above fixed-point system describes the *mutually-consistency condition* characterizing the equilibrium. Notice that the solution is a collection of  $S$  mappings,  $\{\delta_s, \text{ with } s = 1, \dots, S\}$ .

To clarify the framework, the next lines describe two specific examples of incomplete information games.

**Example 1 (Game with Independent types).**  $S = 2$ , each player has linear payoffs and mutually independent information, i.e.  $\pi_s(a_{-s}, x) = \beta_s x - a_s a_{-s}$ ,  $\theta_s \sim F_s$ ,  $\theta_1 \perp \theta_2$ , and  $F_s$  is a spe-

cific distribution function like, say, Normal, Logistic, Uniform, etc. Consider the fixed-point system  $\mu_s(x) = F_s(\beta_s x - \alpha_s \mu_{-s}(x))$  with  $s = 1, 2$ . It represents a system of two equations in the two unknowns  $(\mu_1(x), \mu_2(x))$ . The candidate to be a pure-strategy Bayesian Nash equilibrium is the couple of functions  $\delta_s(x, \theta_s) = 1\{\beta_s x - \alpha_s \mu_{-s}(x) \geq \theta_s\}$ , with  $s = 1, 2$ . To verify this conjecture, we have to prove that these functions satisfy the mutually-consistency condition in Eq. 1. Notice that under the profile  $\delta$ , the probability that player  $-s$  chooses action 1 is  $\sigma_{-s}^\delta(1|x, \theta_1) = \int 1\{\beta_{-s}x - \alpha_{-s}\mu_{-s}(x) \geq \theta_{-s}\} dF_{-s}(\theta_{-s}) = \mu_{-s}(x)$ . Hence,  $1\{\sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma^\delta(a_{-s}|X, \theta_s) - \theta_s \geq 0\} = 1\{\beta_s x - \alpha_s \mu_{-s}(x) \geq \theta_s\} = \delta_s(x, \theta_s)$ , where the first equality follows from the expressions of  $\pi_s$  and  $\sigma^\delta$ , while the last equality follows from the definition of the strategy  $\delta_s$ . From the latter equality, we deduce that the couple of strategies  $\delta_s(x, \theta_s) = 1\{\beta_s x - \alpha_s \mu_{-s}(x) \geq \theta_s\}$ , with  $s = 1, 2$ , satisfies the mutual-consistency condition. So, it represents a Bayesian-Nash equilibrium of the game.

**Example 2 (Game with correlated agents and normal types).** Similarly to the previous example,  $S = 2$ , players' payoffs are linear, but in this case we allow players' types to be mutually correlated. So,  $\pi_s(a_{-s}, \theta_s) = \beta_s x - \alpha_s a_{-s} - \theta_s$ ,  $(\theta_1, \theta_2) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ ,  $\Sigma$  is a 2x2 matrix,  $\text{Var}\theta_1 = \text{Var}\theta_2 = 1$ , and  $\text{Cov}(\theta_1, \theta_2) = \rho$ . Assume that  $\rho \in (-1, 2\pi/(2\pi + |\alpha_1|^2))$ . Denote by  $\Phi$  the cdf of a standard normal, and let  $(\theta_1^*(x), \theta_2^*(x))$  be the solution of the following system of equations

$$\beta_s x - \alpha_s \Phi\left(\frac{\theta_{-s}^*(x) - \rho \cdot \theta_s^*(x)}{\sqrt{1 - \rho^2}}\right) = \theta_s^*(x) \text{ for } s = 1, 2.$$

The candidate to be the pure-strategy Bayesian-Nash equilibrium is the profile  $\delta_s(x, \theta_s) = 1\{\theta_s^*(x) \geq \theta_s\}$  for  $s = 1, 2$ . To verify the mutually consistency condition, notice first that under the profile  $\delta$  the probability of choosing action 1 for player  $-s$  is

$$\sigma^\delta(1|x, \theta_s) = \int 1\{\theta_{-s} \leq \theta_{-s}^*(x)\} \cdot \frac{f_{\theta_{-s}}(\theta_{-s}|\theta_s)}{\Phi((\theta_{-s}^*(x) - \rho\theta_s)/\sqrt{1 - \rho^2})} d\theta_{-s} =$$

so that

$$\begin{aligned} 1\{\sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma^\delta(a_{-s}|X, \theta_s) - \theta_s \geq 0\} = \\ 1\{\beta_s x - \alpha_s \cdot \Phi((\theta_{-s}^*(x) - \rho\theta_s)/\sqrt{1 - \rho^2}) \geq \theta_s\} = \\ 1\{\theta_s \leq \theta_s^*(x)\} = \delta_s(x, \theta_s), \end{aligned}$$

where: the first equality follows from the expressions of  $\sigma^\delta(1|x, \theta_s)$  and  $\pi_s$ ; the second equality follows by noticing that the mapping  $\theta_s \mapsto -\alpha_s \cdot \Phi((\theta_{-s}^*(x) - \rho\theta_s)/\sqrt{1 - \rho^2}) - \theta_s$  is decreasing (as  $\rho \in (-1, 2\pi/(2\pi + |\alpha_1|^2))$ ), while the last equality follows from the definition of  $\delta_s$ . Therefore, the mutually consistency condition is satisfied and  $(\delta_1, \delta_2)$  forms a pure-strategy Bayesian-Nash equilibrium of the game.



The general structure described above is too large. So, following [Liu et al. \(2017\)](#) we consider *monotone pure-strategy equilibria* in those classes of models with *expected payoffs decreasing in the private information*.

**Defnition (Monotone pure strategy equilibrium).** A profile of strategies  $\delta = (\delta_1, \dots, \delta_S)$  is a monotone-decreasing pure-strategy Bayesian-Nash (m.d.p.s. BN) equilibrium for the model described above, if it is an equilibrium -i.e. it satisfies the mutually consistency condition-, and  $\theta_s \succ \delta_s(x, \theta_s)$  is (weakly) decreasing for all  $x$ .

**Assumption M (Monotonically decreasing expected payoffs).** For each profile  $\delta$  of monotone pure strategies, the expected payoff  $\sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma^\delta(a_{-s}|X, \theta_s) - \theta_s$  is continuous and decreasing in  $\theta_s$  for all  $s = 1, \dots, S$ .

As long as we restrict ourselves to consider only monotone pure strategy equilibria, the profile strategies can be written as:

$$\delta_s^*(x, \theta_s) = 1\{\theta_s^*(x) \geq \theta_s\} \text{ for all } s = 1, \dots, S.$$

The monotonicity restriction introduces a separability between the part of the equilibrium strategy depending on the exogenous variables, i.e.  $\theta_s^*(x)$ , and the part which depends on those components unobserved by the econometrician, i.e.  $\theta_s$ . Thanks to this feature, it allows us to characterize the equilibrium strategy using only the collection  $\{\theta_s^*(x) \text{ for } s = 1, \dots, S\}$ . To see this, notice that under the profile  $\delta^* = (\delta_1^*, \dots, \delta_S^*)$ , as long as  $\delta^*$  respects the monotonicity condition -i.e.  $\delta^* = (1\{\theta_1^*(x) \geq \theta_1\}, \dots, 1\{\theta_S^*(x) \geq \theta_S\})$ - the choices of each agent can be characterized by the equivalence

$$\theta_s \leq \theta_s^*(x) \Leftrightarrow \sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma_{\delta^*}(a_{-s}|X, \theta_s) - \theta_s \geq 0 \text{ for all } s = 1, \dots, S.$$

Therefore, using Assumption [M](#), the collection  $\{\theta_s^*(x) \text{ for } s = 1, \dots, S\}$  solves the system

$$\theta_s^*(x) = \sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, x) \cdot \sigma_{\delta^*}(a_{-s}|x, \theta_s^*(x)) \text{ for } s = 1, \dots, S. \quad (2)$$

**Example 1 (continued).** Since the agents' types are mutually independent, the structural part of agent  $s'$  expected payoff,  $\sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma^\delta(a_{-s}|X, \theta_s)$ , will not depend on  $\theta_s$ , hence the monotonicity of the expected payoff will trivially hold for any profile strategy  $\delta$ . Also, the equilibrium  $\delta_s(x, \theta_s) = 1\{\beta_s x - \alpha_s \mu_{-s}(x) \geq \theta_s\}$  (with  $s = 1, 2$ ) is clearly a monotone-decreasing Bayesian-Nash equilibrium.

**Example 2 (continued).** It is evident that since the equilibrium strategies are described by  $\delta_s(x, \theta_s) = 1\{\theta_s^*(x) \geq \theta_s\}$  for  $s = 1, 2$ , the equilibrium is monotone decreasing. For the monotonicity of the expected payoffs, notice that for any profile  $\delta = (1\{\theta_1 \leq \theta_1(x)\}, 1\{\theta_2 \leq \theta_2(x)\})$  of monotone decreasing



strategies, the expected payoff of agent  $s$  is given by

$$\sum_{a_{-s} \in A_{-s}} \pi_s(a_{-s}, X) \cdot \sigma^\delta(a_{-s}|X, \theta_s) - \theta_s = \beta_s x - \alpha_s \Phi((\theta_{-s}(x) - \rho \theta_s) / \sqrt{1 - \rho^2}) - \theta_s$$

for  $s = 1, 2$ . Since  $\rho \in (-1, \sqrt{2\pi/(2\pi + |\alpha_1|^2)})$ , the function on the right-hand side will be decreasing in  $\theta_s$ , so the expected payoff will also be a decreasing function of  $\theta_s$ .

**Rationalization of the Bayesian-Nash Equilibrium.** The fixed-point system in Eq. 2 might admit multiple solutions for some (or all) values of  $x$ . In this section, we restrict our attention to those structures which select a unique equilibrium, while remaining agnostic about the selection mechanism. Such a hypothesis has been used in the literature in empirical games, see e.g. [Bajari et al. \(2010a\)](#), [Brock & Durlauf \(2001\)](#), [Brock & Durlauf \(2007\)](#), and under some conditions will allow us to provide a diagnostic test for the multiplicity of equilibria.

The class of models so far described can be compactly written as

$$\mathbf{M} = \{ (\pi, F_{\theta|X}, \psi) \text{ s.t. } (\pi, F_{\theta|X}) \text{ satisfy E, D, M,} \\ \text{and } \psi \text{ selects a unique m.d.p.s. BN equilibrium} \},$$

where  $\pi = (\pi_1, \dots, \pi_S)$ . Each triple of elements belonging to  $\mathbf{M}$ , say  $(\pi, F_{\theta|X}, \psi)$ , generates a certain distribution (conditional on  $X$ ) of the vector  $a := (a_1, \dots, a_S)$ . Specifically, denote by  $\mathbf{P}$  the set of all distributions of  $a$  conditional on  $X$ . Each distribution in  $\mathbf{P}$  is defined over the measurable space  $(\mathbf{A}, \sigma(\mathbf{A}))$ , where  $\mathbf{A}$  is the space of profile strategies defined at the beginning of this sections, while  $\sigma(\mathbf{A}) = 2^{\mathbf{A}}$  denotes the collection of all the subsets of  $\mathbf{A}$ , i.e. it is the sigma-field generated by  $\mathbf{A}$ . By definition of  $\mathbf{M}$ , for each triple belonging to  $\mathbf{M}$ , say  $(\pi, F_{\theta|X}, \psi)$ , we will have a unique equilibrium  $(\theta_1^*(x), \dots, \theta_S^*(x))$  selected from the solutions of the fixed-point system in Eq. 2. The vector  $(\theta(x), \dots, \theta(x))$  defines a unique set of equilibrium strategies  $a_s = 1\{\theta_s^*(x) \geq \theta_s\}$ , for  $s = 1, \dots, S$ . Therefore, given the cdf of agents' private information,  $F_{\theta|X}$ , we will have a specific distribution of  $a = (a_1, \dots, a_S)$  conditional on the exogenous variables  $X$ . This highlights that the collection  $\mathbf{M}$  generates a mapping  $\mathbf{P}$  which associates to each element of  $\mathbf{M}$  an element of  $\mathbf{P}$ . i.e.  $\mathbf{P} : \mathbf{M} \rightarrow \mathbf{P}$ . Notice that we are not requiring such a mapping to be one-to-one: this would be required for the point identification of the models' fundamentals, but not for our test.

Let us denote with  $\mathbf{P}(\mathbf{M})$  the collection of distributions of  $a = (a_1, \dots, a_S)$  conditional on  $X$  generated by model  $\mathbf{M}$ . So,  $\mathbf{P}(\mathbf{M})$  is the image of  $\mathbf{M}$  through the mapping  $\mathbf{P}$ . Assume that the econometrician observes the variables  $(a, X)$ , and let us denote by  $P_{a|X}$  the *true* distribution of the agents' actions conditional on  $X$ . So,  $P_{a|X}$  is the *de facto* distribution of agents' actions, i.e. it is the *actual* distribution generating the agents' choices that the econometrician wishes to describe. The null hypothesis we want

to test is whether the true distribution of agents' choices,  $P_{a|X}$ , can be generated by the mapping  $P$ , i.e.

$$H_0 : P_{a|X} \in P(M) .$$

From an economic point of view, the above restriction is equivalent to saying that the true distribution of agents' choices can be rationalized, or characterized, by model  $M$ . In other words, when  $H_0$  holds, there exists a structure in  $M$  generating the true probability distribution  $P_{a|X}$ . To build a test for  $H_0$ , we have to characterize the constraints the true distribution  $P_{a|X}$  must satisfy for it to be rationalized by the model  $M$ . Then, testing whether  $P_{a|X}$  satisfies these constraints would be equivalent to testing whether  $P_{a|X}$  can be rationalized by  $M$ . We will therefore use a powerful result obtained by [Liu et al. \(2017\)](#). To display their result, let us denote by  $E_P$  and  $E_P\{\cdot | g(X)\}$  the unconditional and conditional expectation, respectively, computed according to the *true* probability measure  $P$ . Let  $W_s(X) := E_P\{a_s | X\}$ , so  $W_s(X)$  is the actual probability that agent  $s$  makes choice 1 conditional on the exogenous covariates  $X$ . The result from [Liu et al. \(2017\)](#) we use is the following:

**Theorem 1.**  $P_{a|X} \in P(M)$  if and only if for any  $Q = 2, \dots, S$  and for all  $1 \leq s_1 \leq \dots \leq s_Q \leq S$ , the following conditions hold:

- (i)  $E_P\{\prod_{j=1}^Q a_{s_j} | X\} = E_P\{\prod_{j=1}^Q a_{s_j} | W_{s_1}(X), \dots, W_{s_Q}(X)\}$ ;
- (ii)  $E_P\{\prod_{j=1}^Q a_{s_j} | W_{s_1}(X) = \cdot, \dots, W_{s_Q}(X) = \cdot\}$  is strictly increasing and continuously differentiable.

The second part of condition (ii) is only a regularity condition regarding the smoothness of a conditional expectation, something normally assumed in the literature on nonparametric estimation. As [Liu et al. \(2017\)](#) point out, it can be removed by relaxing Assumption [M](#), and not requiring the density of  $\theta$  to be continuously differentiable. The first part of condition (ii) is a bit more stringent but still reasonable, and it would be difficult to provide a counterexample to it. To see this, consider the free-entry game described at the beginning of this section, where each agent is a firm which must decide whether to enter or not to enter a specific market. The monotonicity in condition (ii) requires that if the (conditional) probability of entry of a single firm increases and all the others remain the same, then the joint probability of entry of all the firms must increase as well. Finally, condition (i) requires the conditional expectation  $E\{\prod_{j=1}^Q a_{s_j} | W_{s_1}(X) = \cdot, \dots, W_{s_Q}(X) = \cdot\}$  to be a sufficient statistic for  $E\{\prod_{j=1}^Q a_{s_j} | X = \cdot\}$ . As highlighted by [Liu et al. \(2017\)](#), this seems to be the most stringent one, so we will focus on it in the reminder of the paper.

Before switching to the construction of the test and listing the assumptions needed, it is useful to provide a counterexample to condition (i) of the previous theorem. The one we report here is a refinement of a counterexample provided by [Liu et al. \(2017\)](#).

**Example 3: Non-rationalizable distributions.** Consider the two-players game introduced in Example [1](#), where  $x$  is replaced by  $(x, \eta)$  with  $x = (x_1, x_2)$ , and the profit of each agent writes as

$$\pi_s(x, \eta, a_1, a_2, \theta_s) = a_s \cdot [\psi(x, \eta) - \alpha_s \cdot a_{-s} - \theta_s], \text{ with } \psi(x, \eta) = \phi(x_1) + \eta \cdot \delta(x_2) .$$

Assume that  $\theta_1 \sim U[0, 1], \theta_2 \sim U[0, 1], \theta_1 \perp \theta_2, (\theta_1, \theta_2) \perp x, E\eta = 0, \eta \perp x, x_1 \perp x_2$ . Define  $\mu_s^*(x, \eta)$ , with  $s = 1, 2$ , as the solution of the following system of linear equations

$$\mu_s^*(x, \eta) = \psi(x, \eta) + \alpha_s \cdot \mu_{-s}^*(x, \eta) \text{ for } s = 1, 2.$$

If  $\alpha_1 \cdot \alpha_2 \neq 1$ , the solution is simply represented by

$$\mu_s^*(x, \eta) = \frac{1 + \alpha_s}{1 - \alpha_1 \cdot \alpha_2} \cdot \psi(x, \eta) \text{ with } s = 1, 2.$$

Assume that for all  $(x, \eta)$  in the support it holds that  $\psi(x, \eta) + \alpha_s \cdot \mu_{-s}^*(x, \eta) \in [0, 1]$ , i.e.  $\frac{1 + \alpha_s}{1 - \alpha_1 \cdot \alpha_2} \cdot \psi(x, \eta) \in [0, 1]$ . Hence, denoting with  $F_{\theta_s}$  the cdf of  $\theta_s$ , the couple  $(\mu_1^*, \mu_2^*)$  satisfies

$$\mu_s^*(x, \eta) = F_{\theta_s}(\psi(x, \eta) + \alpha_s \cdot \mu_{-s}^*(x, \eta)) \text{ for } s = 1, 2.$$

By the above display and following the same steps as in Example 1, the Bayesian-Nash equilibrium of the game is represented by

$$a_s^* = \{ \psi(x, \eta) + \alpha_s \cdot \mu_{-s}^*(x, \eta) \geq \theta_s \} \text{ for } s = 1, 2.$$

From the equilibrium strategies and the definition of  $(\mu_1^*, \mu_2^*)$ ,

$$E\{a_s^* | x, \eta\} = \frac{1 + \alpha_s}{1 - \alpha_1 \cdot \alpha_2} \cdot \psi(x, \eta) \text{ for } s = 1, 2, \text{ and } E\{a_1^* \cdot a_2^* | x, \eta\} = E\{a_1^* | x, \eta\} \cdot E\{a_2^* | x, \eta\}.$$

Now, assume that the data  $(a_1^*, a_2^*)$  is generated by the above game, and that  $\eta$  is an unobserved heterogeneity term. In particular, both players observe  $(x, \eta)$ , while the researcher does not observe  $\eta$ . Since the researcher only observes  $(a_1^*, a_2^*, x)$ , her goal is to test whether the distribution of such variables can be rationalized by a unique Bayesian-Nash equilibrium of a game that respects Assumptions E, D, M, where  $x$  is the only exogenous covariate. This holds only if  $E\{a_1^* \cdot a_2^* | x\} = E\{a_1^* \cdot a_2^* | W_0(x)\}$ . However, such an equality does not hold for the example considered here. In fact,

$$E\{a_s^* | x\} = \frac{1 + \alpha_s}{1 - \alpha_1 \cdot \alpha_2} \cdot \phi(x_1) \text{ for } s = 1, 2,$$

so by using  $E\eta = 0, x_1 \perp x_2$ , and  $\eta \perp x$ , we get

$$E\{a_1^* \cdot a_2^* | x\} = E\{a_1^* | x\} \cdot E\{a_2^* | x\} + \frac{1 + \alpha_1}{1 - \alpha_1 \cdot \alpha_2} \cdot \frac{1 + \alpha_1}{1 - \alpha_1 \cdot \alpha_2} \cdot \delta(x_2)^2 \cdot E\eta^2$$

Define  $W_{0,s}(x) := E\{a_s^* | x\}$  and  $W_0(x) := (W_{0,1}(x), W_{0,2}(x))$ . Applying the conditional expectation

operator  $E\{\cdot | W_0(x)\}$  to both sides of the above display,

$$E\{a_1^* \cdot a_2^* | W_0(x)\} = W_{0,1}(x) \cdot W_{0,2}(x) + \frac{1+\alpha_1}{1-\alpha_1 \cdot \alpha_2} \cdot \frac{1+\alpha_1}{1-\alpha_1 \cdot \alpha_2} \cdot E\{\delta(x_2)^2\} \cdot E\{\eta^2\},$$

where we have used the Law of Iterated Expectations,  $W_{0,s}(x) = E\{a_s^* | x\} = \frac{1+\alpha_s}{1-\alpha_1 \cdot \alpha_2} \cdot \phi(x_1)$ ,  $\eta \perp x_1$ , and  $x_1 \perp x_2$ . So, subtracting the latter display to the former delivers

$$E\{a_1^* \cdot a_2^* | x\} - E\{a_1^* \cdot a_2^* | W_0(x)\} = \frac{(1+\alpha_1) \cdot (1+\alpha_2)}{(1-\alpha_1 \cdot \alpha_2)^2} \cdot E\{\eta^2\} \cdot \delta(x_2)^2 - E\{\delta(x_2)^2\} \cdot \sum$$

which is different from zero as long as  $\delta(x_2)$  is a non-degenerate function. Since  $E\{a_1^* \cdot a_2^* | X\} \neq E\{a_1^* \cdot a_2^* | W_0(X)\}$ , the observed data  $(a_1^*, a_2^*, X)$  generated by the above game-theoretical model has a distribution that cannot be rationalized by a Bayesian-Nash equilibrium of a game respecting Assumptions [E](#), [D](#), [M](#). In other words, since the researcher is omitting an unobserved variable  $\eta$ , any game-theoretical model assuming that the data  $(a_1^*, a_2^*, x)$  is generated by a Bayesian-Nash equilibrium respecting Assumptions [E](#), [D](#), [M](#) will be misspecified.

### 3 The Test

For simplicity of exposition, we focus on a 2-players game. At the end of this section, we will provide the general form of the test with any finite number of players. Let us assume to observe a cross-section of agents' decisions,  $(a_1, a_2)$ , and the exogenous covariates,  $X \in \mathbb{R}^p$ . We do not assume *a priori* that these observations are generated by an incomplete-information game played multiple times, as this is exactly the restriction we want to test. As an example, such a data set might consist of airline companies decisions about offering a connection between two different airport hubs, as in [Ciliberto & Tamer \(2009\)](#). Define  $W_{0,1}(X) := E\{a_1 | X\}$ ,  $W_{0,2}(X) := E\{a_2 | X\}$ . Here we have dropped the index  $P$  to the expectation  $E$  with respect to the previous section, for the ease of notation. Let  $W_0(X) := (W_{0,1}(X), W_{0,2}(X))$ , i.e.  $W_0(X)$  stands for the vector of the conditional probabilities that agents 1 and 2 take decision 1, respectively. For example, in the data set considered in [Ciliberto & Tamer \(2009\)](#),  $W_{0,1}(X)$  denotes the probability, conditional on  $X$ , that firm 1 offers a connection between two different cities (or airports). By the discussion provided in the previous section, to test whether the distribution of the observed data  $P_{a|X}$  can be rationalized by a single m.d.p.s. BN equilibrium, we must test condition (ii) of Theorem 1. With the present notation it writes as

$$H_0 : E\{a_1 \cdot a_2 | X\} = E\{a_1 \cdot a_2 | W_0(X)\}.$$

Denote with  $f_{W_0}$  the density of  $W_0(X)$  with respect to the Lebesgue measure. We introduce  $Y := a_1 \cdot a_2$ ,  $m_{W_0}(w) := E\{Y | W_0(X) = w\}$ , and  $\varepsilon := Y - m_{W_0}(W_0(X))$ . The null hypothesis written above is

equivalent to

$$H_0 : E\{\varepsilon \cdot f_{W_0}(W_0(X)) \mid X\} = 0.$$

To build a test for  $H_0$ , we transform the above *conditional* moment restriction into a *continuum* of *unconditional* moment restrictions. So, let  $T$  be a compact subset of  $\mathbb{R}^p$  encompassing the origin, with  $p = \dim(X)$ . For  $t \in T$ ,  $\varphi_t(\cdot) := \phi(t^T \cdot)$ , with  $\phi$  being an analytic non-polynomial function defined on  $\mathbb{R}$ , i.e.  $\phi$  is a one-variable function infinitely times continuously differentiable which does not have a polynomial form. Examples are  $\exp(\cdot)$ ,  $\cos(\cdot)$ ,  $\sin(\cdot)$ ,  $\exp(i \cdot)$  -where  $i = \sqrt{-1}$ . By the results in [Bierens & Ploberger \(1997\)](#) and [Stinchcombe & White \(1998\)](#), the null hypothesis  $H_0$  is equivalent to

$$H_0 : E\{\varepsilon \cdot f_{W_0}(W_0(X)) \cdot \varphi_t(X)\} = 0 \text{ for all } t \in T.$$

We test  $H_0$  versus its logical complement  $H_1 : H_0$ . To simplify the notational burden, define the empirical-mean operator

$$P_n := \frac{1}{n} \sum_{i=1}^n \delta_{Z_i} \text{ with } Z_i := (a_{1,i}, X_{2,i})$$

For any function  $g$  of  $Z$ , we have  $P_n g = (1/n) \cdot \sum_{i=1}^n g(Z_i)$ . We also define  $Pg := \int g(z) P(dz)$ . Notice that if  $g$  is a non-random function,  $Pg = E\{g(Z)\}$ . With this notation, the null hypothesis can be re-formulated as

$$H_0 : P \varepsilon f_{W_0} \varphi_t = 0 \quad \forall t \in T.$$

A feasible statistic for the above condition is

$$S_n = \int \left\| \frac{1}{\sqrt{n}} \cdot P_n \varepsilon \hat{f}_{W_0} \varphi_t \right\|^2 \mu(dt),$$

where  $\|\cdot\|$  stands for a norm on  $\mathbb{C}$ ,  $(\varepsilon, \hat{f}_{W_0})$  is an estimator of  $(\varepsilon, f_{W_0})$ , and  $\mu$  is a measure absolutely continuous with respect to the Lebesgue measure. Intuitively, if the null hypothesis holds true, one would expect that  $\frac{1}{\sqrt{n}} \cdot P_n \varepsilon \hat{f}_{W_0} \varphi_t$  is bounded in probability and converges to a specific distribution, so that also  $S_n$  will converge to a tight distribution. Differently, under the alternative hypothesis  $H_1$ ,  $\frac{1}{\sqrt{n}} \cdot P_n \varepsilon \hat{f}_{W_0} \varphi_t$  will explode and also the statistic  $S$  would diverge.

*Remark 2.* The statistic we are proposing is an Integrated Conditional Moment (ICM) Test. Since our proofs hold for any continuous functional defined on  $A^\infty(T)$ , other tests would also be possible<sup>1</sup>. For instance, we could consider  $\sup_{t \in T} \left| \frac{1}{\sqrt{n}} \cdot P_n \varepsilon \hat{f}_{W_0} \varphi_t \right|$ , a Kolmogorov-Smirnov (KS) statistic. However, under a suitable choice of  $\varphi$ , the ICM statistic is much easier to compute than the KS test, as we will show in the next section.

<sup>1</sup> $A^\infty(T) := \{g : T \rightarrow \mathbb{R} \text{ such that } \sup_{t \in T} |g(t)| < \infty\}$

*Remark 3.* The above statistic is used to test the null hypothesis within a two-players game. Differently, in the presence of  $S$  different players there will be  $2^S - (S + 1)$  moment conditions to test. In this case, we proceed in the same way as above, and compute for each moment equality  $c$  the corresponding statistic  $S_{n,c}$ .  $H_0$  can then be tested through

$$S_n = \sum_c S_{n,c} \text{ or } S_n = \max_c S_{n,c}.$$

Since both the “sum” and the “max” operator are continuous transformations, our proofs also hold for such functionals.

To compute the statistic  $S_n$  we need to provide a feasible estimator of  $(\varepsilon, f_{W_0})$ . Notice that we have a problem of non-observability of the regressors  $\{W_0(X_i)\}_{i=1}^n$ . Because of its non-observability,  $W_0(X)$  is said to be a generated regressor. Due to this feature, it is natural to proceed to a two-step estimation: in a first step, we estimate nonparametrically  $W_0$ ; in a second step, we replace the generated regressors with their estimates,  $\{\hat{W}(X_i)\}_{i=1}^n$ , and proceed to the nonparametric estimation of both  $m_{W_0}$  and  $f_{W_0}$ . Both these steps involve nonparametric estimators, and it is well known that nonparametric methods provide biased estimates. In our case, the bias might have a relevant magnitude, as the second step estimate will depend on the fits of the first step estimator. The bias will impact negatively on the capacity of  $(\hat{W}, \hat{m}_{\hat{W}})$  to match their respective targets, so it will also impact negatively on the performance of the ICM test. To shed more light on this aspect, imagine to use a kernel method for  $\hat{m}_{\hat{W}}$  and  $\hat{W}$ , so that such estimate will be based on a kernel function, say  $K$ , and a bandwidth sequence, say  $h^2$ . To deal with the bias showing up in  $S_n$  -which arises from the nonparametric estimation of  $m_{W_0}$  and  $W_0$ - the approach usually taken in the literature on nonparametric testing is based on two features. First, a high-order kernel is used, so the kernel  $K$  is set to be a relatively irregular function. Second, the bandwidth rate is set according to a method that undersmooths, and hence restricts the set of bandwidths admissible for the test. For example, the bandwidth that minimizes the Mean Squared Error is not admissible in the presence of undersmoothing. Technically, the bandwidth  $h$  and the kernel  $K$  must be such that  $n \cdot h^{2r} = o(1)$ , where  $r$  is the order of the kernel -i.e. the larger  $r$  the more  $K$  will be an irregular function. Such a condition will have a negative impact on the bias-variance trade-off of the nonparametric estimator. Assume that for a given bandwidth sequence  $h$ , the order  $r$  is set to be larger than 4 to respect the condition  $nh^{2r} = o(1)$ . Accordingly, the kernel will be a relatively irregular function and will inflate the variance. To limit such a variance inflation, the usual approach in the literature is to ignore, to a certain extent, the condition  $nh^{2r} = o(1)$  and hence to choose a kernel of low order that, however, will inflate the bias term. In other words, the bias arising from the nonparametric estimation will impact negatively on the capacity of the estimators to fit well the original functions, and hence on the capacity of the test to provide the good size and power.

<sup>2</sup>In this context, since we have two nonparametric functions to estimate,  $(m_{W_0}, W_0)$ , we should also introduce two kernels  $(K_1, K_2)$  and two bandwidth rates  $(h_1, h_2)$ . This aspect, however, is not essential for the present discussion.

---

**Algorithm 1** General Boosting method for regressions

---

1: Initialize with an estimate  $m^{[0]}$  of the regression  $m$

2: Increase  $b$  by 1, i.e.  $b \leftarrow b + 1$ ; compute the residuals  $\varepsilon^{[b]} = Y - m^{[b-1]}$  and hence the sample  $\{\varepsilon_i^{[b]}, W_i\}_{i=1}^n$ .

3: Fit the residuals  $\{\varepsilon_i^{[b]}\}_{i=1}^n$  to the explanatory variables  $\{W_i\}_{i=1}^n$ , according to the base estimation procedure, and hence obtain the fitted function  $r^{[b]}$

4: Update the estimator as  $m^{[b]} = m^{[b-1]} + r^{[b]}$

5: repeat Step 2 to Step 4  $B$  times

---

## 4 Estimation by L2 Boosting

To deal with the bias arising from the nonparametric estimation, a method that has been proven to be quite effective is *Boosting*. This is an algorithm originally conceived in the Machine Learning literature which gained a strong success in the statistical literature on estimation, see ?, [Di Marzio & Taylor \(2008\)](#), [Park et al. \(2009\)](#), [Cornillon et al. \(2014\)](#). The main idea is to estimate iteratively an object of interest on different samples, and then aggregate these estimates in a unique estimator. Imagine to estimate a regression function  $m$ , where the response variable is  $Y$  and the set of regressors is  $W$ . A common Boosting algorithm for regressions is based on three elements: (i) an initial estimator; (ii) a base estimation procedure; (iii) the number of iterations to perform. It is described in Algorithm 1.

At each iteration, the procedure applies the base estimation method to the residuals obtained from the previous iteration. Different base methods give rise to different boosting algorithms. The specific boosting algorithm we use in the present work is the  $L_2$ -boosting. This uses a base procedure which minimizes a least-squares criterion, see ? and [Di Marzio & Taylor \(2008\)](#). To apply the  $L_2$ -boosting procedure to the computation of  $S_n$ , recall that our estimation is based on two steps. In a first step, we estimate the generated regressors -i.e. the function  $W_0$ -, while in a second step we proceed to the estimation of  $m_{W_0}$  and  $f_{W_0}$  by replacing the unobserved regressors  $\{W_0(X_i)\}_{i=1}^n$  with their estimated counterpart  $\{\hat{W}(X_i)\}_{i=1}^n$ . So, let us assume to have an estimator  $\hat{W}$  of the generated regressor  $W_0$ . Let  $\hat{m}_W^{[0]}$  be an initial estimator, i.e. a *weak learner*, of  $m_{W_0}$ . The  $L_2$ -boosting algorithm applied to the estimation of  $m_{W_0}$  is described in Algorithm 2. After  $B$  boosting iterations, the initial weak learner  $\hat{m}_W^{[0]}$  is transformed into the *deep learner*  $\hat{m}_W^{[B]}$ . The boosting estimator after  $b$  iterations can be written as

$$\hat{m}_W^{[b]} = \hat{m}_W^{[b-1]} + \frac{\hat{\varepsilon}_{b-1}}{\hat{f}_W^{[0]}} \quad \hat{T}_{\hat{\varepsilon}_{b-1}} = \frac{1}{nh^d} \sum_{i=1}^n \hat{\varepsilon}_i^{[b-1]} \hat{f}_{n,i} K \cdot \frac{\hat{W}_i}{h}, \quad (3)$$

for  $b = 1, \dots, B$ , where  $K$  is a kernel function,  $\hat{\varepsilon}_i^{[b-1]} = Y_i - \hat{m}_W^{[b-1]}(\hat{W}_i)$ , and  $\hat{f}_W^{[0]}$  is an estimator of  $f_{W_0}$ .



---

**Algorithm 2**  $L_2$  Boosting
 

---

- Initialize with estimates  $\hat{f}_{\hat{W}}^{[0]}$  and  $\hat{m}_{\hat{W}}^{[0]}$
- Increase  $b$  by one, i.e.  $b \leftarrow b + 1$ , and compute the residuals  $\varepsilon_{\hat{W}}^{[b-1]} = Y - \hat{m}_{\hat{W}}^{[b-1]}$
- From the sample  $\{\varepsilon_i^{[b-1]}, \hat{W}_i\}_{i=1}^n$  compute  $\hat{T}_{\varepsilon_{b-1}}^{\hat{W}}$  (defined in Eq. 3) and update the estimator of  $m_{W_0}$  as

$$\hat{m}_{\hat{W}}^{[b]} = \hat{m}_{W_0}^{[b-1]} + \frac{\hat{T}_{\varepsilon_{b-1}}^{\hat{W}}}{\hat{f}_{W_0}^{[0]}}$$

Repeat the above steps  $B$  times.

---

$\hat{t}_{ni}$  is a trimming factor useful to handle a random denominator,

$$\hat{t}_{ni} := \mathbf{1}' \hat{f}(X_i) \geq \tau_n', \quad t_{n,i} := \mathbf{1}' f(X_i) \geq \tau_n',$$

with  $f$  density of  $X$ .

*Remark 4.* When  $\hat{f}_{\hat{W}}^{[0]}$  and  $\hat{m}_{\hat{W}}^{[0]}$  are kernel estimators built with the same kernels and bandwidths as  $\hat{T}_{\varepsilon_b}$  the boosting described in Algorithm 2 is the same as in [Di Marzio & Taylor \(2008\)](#). Here, we are not specifying the initial estimators  $\hat{f}_{\hat{W}}^{[0]}$  and  $\hat{m}_{\hat{W}}^{[0]}$  for the aim of generality of the theory and simplicity of proofs. Notice that the density  $\hat{f}_{\hat{W}}^{[0]}$  is not updated in the boosting algorithm of the regression, see [Di Marzio & Taylor \(2008\)](#).

*Remark 5.* We have chosen to keep  $\hat{W}$  in a general form and not to specify a particular method used in the first-step.  $\hat{W}$  will only have to respect some high-level conditions introduced in the next section. This allows us to keep a good level of generality and to build clean proofs in an acceptable number of pages. However,  $\hat{W}$  can be computed according to the same boosting procedure as in Algorithm 2.

The procedure just described delivers the estimators  $\varepsilon_{\hat{B}}$  of the residual  $\varepsilon$ . From this, we construct the test statistic  $S_n$  based on  $B$  boosting iterations as

$$S_n := \int \sqrt{n} \mathbf{P}_n \varepsilon_{\hat{B}} \hat{f}_{\hat{W}}^{[0]} \varphi_t |^2 \mu(dt).$$

If we choose the weighting function  $\varphi$  to be the complex exponential  $\exp(i \cdot)$ , with  $i = \sqrt{-1}$ , the statistic  $S_n$  will have the following simple form

$$S_n = \frac{1}{n} \sum_{i,j} \varepsilon_{\hat{B},i} \cdot \hat{f}_{\hat{W},i}^{[0]} \cdot \varepsilon_{\hat{B},j} \cdot \hat{f}_{\hat{W},j}^{[0]} \cdot \phi_{\mu}(X_i - X_j),$$



where  $\phi_\mu$  is the characteristic function of the measure  $\mu$ , and  $\hat{f}^{[0]} := \hat{f}^{[0]}(W_{i_W})$ . For instance, if  $\mu$  is chosen to be the multivariate standard normal distribution, then  $\phi_\mu(x)$  boils down to  $\prod_{j=1}^p \exp(x_j)$ , where  $x_j$  is the  $j$ -th component of  $x$ .

## 5 The Assumptions and the Asymptotic Test

Let us recall that  $a := (a_1, a_2)$  and  $W_0 := (W_{0,1}, W_{0,2})$ . In what follows,  $d$  denotes the number of components of  $W_0(X)$  and  $\hat{W}(X)$ . In a two-players context,  $d = 2$ . Having a general  $d$  allows to identify better the role of the dimension of  $W_0$  on the assumptions and the bandwidths. This in turn clarifies the extension to a generic number of players. Define the sets

$$W_n := \{w : f_{W_0}(w) > \frac{\tau_n}{2}\} \text{ and } X_n := \{x : f(x) > \frac{\tau_n}{2}\}. \quad (4)$$

For a vector of positive natural numbers  $k = (k_1, \dots, k_d)$ , define the differential operator<sup>3</sup>

$$\partial^k := \frac{\partial^k}{\partial^{k_1} w_1 \dots \partial^{k_d} w_d},$$

with  $k = k_1 + \dots + k_d$ . Given a generic real set  $D$ , with  $D$  convex, the class of smooth functions we consider is

$$C(D) := \{g : D \rightarrow \mathbb{R} : \max_{k \leq \frac{\tilde{d}^+}{2} + 1} \|\partial^k g\|_{\infty, D} \leq M\}, \quad (5)$$

with  $\tilde{d}^+$  the largest even number weakly smaller than  $\tilde{d} := \dim(D)$ . If  $\tilde{d}$  is even then  $\tilde{d}^+ = \tilde{d}$ ; if  $\tilde{d}$  is odd then  $\tilde{d}^+ = \tilde{d} - 1$ .

**Observation.**  $C(D) \subset C_M^\alpha(D)$  for  $\alpha = \frac{d^+}{2} + 1$ , where  $C_M^\alpha(D)$  is the class of functions defined at page 154 in [van der Vaart & Wellner \(1996\)](#).

Let  $K_\lambda^r$  be the class of functions  $(v_1, \dots, v_d) \mapsto \prod_{j=1}^d k(v_j)$  with  $k$  univariate kernel of order  $r$  that is  $\lambda$  times continuously differentiable. Denote  $p_n = P(f(X) \leq \tau_n^2)$  and  $C_n = C(W_n)$ .

**Assumption 1.**  $\{Y_i, X_i\}_{i=1}^n$  is a sequence of iid and bounded random variables.

**Assumption 2.** (i)  $m_{W_0}, f_{W_0} \in C(W_0)$  and  $W \in C(X)$ ; (ii)  $X_\infty$  admits a density conditionally on  $W_0(X)$ , denoted by  $f_{X|W_0(X)}$ ; (iii)  $m_{W_0}, f_{W_0}$ , and  $f_{X|W_0(X)}(\cdot | \cdot)$  are  $rB$  times continuously differentiable with uniformly bounded derivatives.

**Assumption 3.**  $K \in K_\lambda^r$ , and  $\lambda = \frac{d^+}{2} + 1$ .

<sup>3</sup>We are using a similar notation as in [van der Vaart & Wellner \(1996\)](#).

**Assumption 4.** (i)  $p_n n^{1/2} = o(1)$ ,  $\frac{p_n n^{1/4}}{h^{d_{\tau_n^B}}} = o(1)$ , and for each  $n$  large enough  $W_n$  and  $X_n$  are convex; (ii) there exists  $\eta$  small enough and  $N$  large enough such that for  $n \geq N$  if  $f(x) \geq \tau_n$  then  $f_{W_0}(W_0(x)) \geq \eta \tau_n$ ; (iii)  $\frac{h}{\tau_n} = o(1)$ .

**Assumption 5.**  $\frac{\log n}{n^{1/2} h^{d_{\tau_n^B}}} = o(1)$ ,  $n h^{4r_B} = o(1)$ .

**Assumption 6.** (i) for  $b_n \in \{\hat{t}_n, t_n\}$ ,  $\|(\hat{W} - W_0)b_n\|_\infty = o_P(\tau_n n^{-1/4})$ ; (ii)  $P(\hat{W} \in \mathcal{C}(X_n)) \rightarrow 1$ ; (iii) For any class of functions  $\Psi$  satisfying condition [CL](#) (see Appendix [A](#)) and for  $\phi_t(x) := E\{\psi_t(Z) | X = x\}$ ,

$$\sqrt{n} P_n(\hat{W} - W_0) t_n \psi_t = \sqrt{n} P_n(a - W_0) \phi_t + o_P(1) \text{ uniformly over } \mathcal{T}.$$

**Assumption 7.** (i) Rates for the weak learners: for  $b_n \in \{\hat{t}_n, t_n\}$ ,  $\|(\hat{m}_W^{[0]} - m_{W_0})b_n\|_\infty = o_P(d_n)$ ,  $\|(\hat{\partial m}_W^{[0]} - \partial m_{W_0})b_n\|_\infty = o_P(d_n)$ ,  $\|(\hat{f}_W^{[0]} - f_{W_0})b_n\|_\infty = o_P(d)$ ,  $\|(\hat{\partial f}_W^{[0]} - \partial f_{W_0})b_n\|_\infty = o_P(d)$ , with  $\frac{d_n}{\tau_n} = o(n^{-1/4})$ ;

(ii) Regularity of the weak learner:  $P(\hat{m}_W^{[0]} \in \mathcal{C}_n) \rightarrow 1$ ;

(iii) Regularity of the boosting updates:  $P(\frac{\hat{T}_{\varepsilon_b}}{f_{W_0}} \in \mathcal{C}_n \rightarrow 1 \text{ for } b = 1, \dots, B-1)$ .

**Comments on the assumptions.** Assumptions [1-3](#) are standard in the literature on nonparametric estimation and testing. Assumption [4](#) is needed because of the generated variables. It essentially imposes that the tails of  $X$  do not have to be too thick. Several versions of it can be found in e.g. [Escanciano et al. \(2014\)](#) and [Escanciano et al. \(2016\)](#). Assumption [5](#) establishes that the kernel order and the bandwidth used in the boosting updates (i.e. for  $T$ ) must be decided in connection with the number of boosting iterations  $B$ . Assumption [6](#) is a high-level condition for the first-stage estimator of the conditional probabilities. In particular, Assumption [6](#)(i) establishes the convergence rates of  $\hat{W}$ . Such convergence rates are relatively standard in the literature and similar to e.g. [Escanciano et al. \(2014\)](#), [Mammen et al. \(2012\)](#), and [Andrews \(1995\)](#). Assumption [6](#)(ii) requires that the first-stage estimator must belong to a “regular” class of functions wpt1. This is also a general regularity condition often assumed in the literature, see [Escanciano et al. \(2014\)](#), [Mammen et al. \(2012\)](#), and [Andrews \(1995\)](#). Assumption [6](#)(iii) requires an expansion for the first-stage estimator, and it is similar in nature to Assumption 10 in [Escanciano et al. \(2014\)](#). Both Assumptions [6](#)(i) and [6](#)(ii) can be proved using results contained in, e.g. [Escanciano et al. \(2014\)](#), [Li & Racine \(2006\)](#), or [Andrews \(1995\)](#). The expansion in Assumption [6](#)(iii) can be obtained by specifying the first-step estimator for the generated regressors. Then, we could proceed similarly as in the proofs of Appendix [A](#), or using the results in [Escanciano et al. \(2014\)](#). Assumption [7](#) has the same spirit as Assumption [6](#) and refers to the second-stage *weak learners* and *deep learners*. The main difference is that Assumption [7](#) is imposing a  $n^{-1/4}$  convergence rate on the first-order derivatives  $\hat{\partial n}_W^{[0]}$  and  $\hat{\partial f}_W^{[0]}$ . Such rates are used to deal with the estimation error coming from the first-stage (i.e. from  $\hat{W}$ ). These rates will be used in connection with a first-order Taylor expansion. Similar rates are

---

**Algorithm 3** Multinomial Bootstrap DGP

---

- 1: Resample  $a^*$  according to a Bernoulli distribution with probability of success  $\hat{W}_{1,i}^{[B_1]}$ , i.e.  $a_{1,i}^* \sim \text{Bernoulli}(\hat{W}_{1,i}^{[B_1]})$
- 2a: If  $a_{1,i}^* = 1$ ,  $a_{2,i}^* \sim \text{Bernoulli} \left( \frac{m_{\hat{W}}^{[B_1]}(\hat{W}_i)}{\hat{W}_i^{[B_1]}} \right)$
- 2b: If  $a_{1,i}^* = 0$ ,  $a_{2,i}^* \sim \text{Bernoulli} \left( \frac{\hat{W}_{2,i} - m_{\hat{W}}^{[B_1]}(\hat{W}_i)}{1 - \hat{W}_{1,i}} \right)$
- 3: Repeat Step 1 - Step 2  $n$  times and save the sample  $\{a_{1,i}^*, a_{2,i}^*\}_{i=1}^n$
- 

also assumed in [Rothe \(2009\)](#). The high-level conditions contained in Assumption [6](#) and [7](#) are useful for providing a clean exposition. They dramatically simplify the proofs and allow us to contain such proofs in an acceptable number of pages.

The asymptotic distribution of our statistic is obtained in the following

**Proposition 6.** *Under Assumptions [1-7](#) if  $H_0$  holds*

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}) \varphi_t = \sqrt{n} \mathbf{P}_n \varepsilon f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n (a - W_0)^T \partial m_{W_0} f_{W_0} \varphi_t^\perp + o_p(1)$$

uniformly over  $\mathbf{T}$ , where  $\varphi_t^\perp(X) := \varphi_t(X) - \iota_t(W_0(X))$  and  $\iota_t(W_0(X)) := E\{\varphi_t(X) \mid W_0(X)\}$ . Accordingly,

$$S_n = \int \left( \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}) \varphi_t \right)^2 \mu(dt) + \int \mathbf{G} \dot{\mu}(dt),$$

where  $\mathbf{G}$  is a Gaussian stochastic process taking values in  $A(\mathbf{G})$  and defined by the collection of covariances  $P\varepsilon^2 \phi_{t_1 t_2}^\perp: t_1, t_2 \in \mathbf{T}$ , with  $\mathbf{G} := (a, y, x) \mapsto (y - m_{W_0}(W_0(x))) \cdot f_{W_0}(W_0(x)) \cdot \varphi_t^\perp(x) - (a - W_0(x)) \partial m_{W_0}(W_0(x)) f_{W_0}(W_0(x)) \varphi_t^\perp(x) : t \in \mathbf{T}$ .

## 6 Multinomial Bootstrap

Proposition [6](#) shows that the asymptotic null distribution of  $S_n$  depends on unknown features of the data. So, the asymptotic null distribution cannot be used to approximate the critical values of the test. We therefore propose a novel *Multinomial Bootstrap* procedure to simulate the critical values. Such procedure resamples the data using the information contained in  $H_0$ . In particular, it assumes that the data is generated by a distribution that can be rationalized by a unique Bayesian-Nash equilibrium of a binary game. Since the information under  $H_0$  is exploited in the bootstrap scheme, we should obtain good properties in terms of size and power:

Denote with  $\text{Bernoulli}(p)$  a Bernoulli distribution with probability of success equal to  $p$ . The Multinomial Bootstrap resamples the observations on  $(a_1, a_2)$  according to the procedure summarized in Algorithm 3. Such artificial DGP generates the sample  $\{a_{1,i}^*, a_{2,i}^*, X_i\}_{i=1}^n$ . By replacing this sample to the original one, i.e. to  $\{a_{1,i}, a_{2,i}, X_i\}_{i=1}^n$ , and then implementing the same procedure as in Section 3 -including the  $L_2$  boosting iterations-, we obtain the bootstrap versions  $(\hat{\varepsilon}_B^*, S_n^*)$  of  $(\hat{\varepsilon}_B, S_n)$  :

$$\hat{\varepsilon}_B^* := a_{1,2}^{**} - m_{\hat{W}^*}^{[B]*}(\hat{W}^*), \quad S_n^* = \int \sqrt{\frac{1}{n}} \cdot \mathbf{P}_{\hat{W}^*}^{**} \hat{f}_{\hat{W}^*}^{[0]*} | \mu(dt),$$

where  $\mathbf{P}_{\hat{W}^*}^* = \frac{1}{n} \sum \delta_{Z_i^*}$  and  $Z_i^* := (a_{1,i}^*, a_{2,i}^*, X_i)$ .

The Multinomial Bootstrap just described has two attractive features. First, it preserves the 0-1 nature of the variables  $a_1^*$  and  $a_2^*$ . Second, it implements *all* the constraints on the joint distribution of  $(a_1, a_2)$  which are suggested by the null hypothesis. Denote with  $P_i^*$  the bootstrap probability measure of  $(a_1^*, a_2^*)$  conditional on  $X_i$  defined in Step 1-2 of Algorithm 3. Notice that

$$P_i^*(a_1^* = 1, a_2^* = 1 | X_i) = m_{\hat{W}}^{[B]}(\hat{W}_i), \quad P_i^*(a_1^* = 1, a_2^* = 0 | X_i) = \hat{W}_{1,i} - m_{\hat{W}}^{[B]}(\hat{W}_i),$$

$$P_i^*(a_1^* = 0, a_2^* = 1 | X_i) = \hat{W}_{2,i} - m_{\hat{W}}^{[B]}(\hat{W}_i).$$

From the above display, the artificial DGP specified in Algorithm 3 generates a probability  $P_i^*$  that equals, up to an estimation error, the probability  $P_{(a_1, a_2) | X_i}$  when  $H_0$  holds true. To see this, denote with  $P_{(a_1, a_2) | X}^{H_0}$  the probability distribution  $P_{(a_1, a_2) | X}$  under  $H_0$ . Then,

$$P_{(a_1, a_2) | X}^{H_0}(a_1 = 1, a_2 = 1 | X) = E\{a_1 a_2 | W(X)\}, \quad P_{(a_1, a_2) | X}^{H_0}(a_1 = 1, a_2 = 0 | X) = W_1(X) - E\{a_1 a_2 | W(X)\}$$

$$P_{(a_1, a_2) | X}^{H_0}(a_1 = 0, a_2 = 1 | X) = E\{a_2 \cdot (1 - a_1) | W(X)\} = W_{0,2}(X) - E\{a_1 a_2 | X\}.$$

The last two displays highlight that  $P_i^*$  is the same, up to an estimation error, as the probability  $P_{(a_1, a_2) | X_i}^{H_0}$ . In this sense, the Multinomial Bootstrap exploits *all* the constraints imposed by  $H_0$  on the distribution of  $(a_1, a_2)$ .

To show the validity of the bootstrap scheme, we need some adding assumptions. Denote with  $P^*$  the probability measure that considers only the data  $\{a_{1,i}^*, a_{2,i}^*\}_{i=1}^n$  as random and assumes as fixed the sample data  $\{a_{1,i}, a_{2,i}, X_i\}_{i=1}^n$ . For a random variable  $Z$  measurable with respect to  $\{a_{1,i}^*, a_{2,i}^*, a_{1,i}, a_{2,i}, X_i\}_{i=1}^n$ , the notation  $Z_n = o_P(1)$  means that  $P(|Z| > \delta) \xrightarrow{P} 0$  for every  $\delta > 0$ . Let  $\hat{W}^*$  be the bootstrap counterpart of  $\hat{W}$ .

**Assumption 8.** There exists an  $s \in (0, 1)$  such that  $m_{W_0}(W_0(X)), W_{0,j}(X) \in (s, 1 - s)$  for  $j = 1, 2$ .

**Assumption 9.** (i) for  $b_n \in \{t_n, t_n^*\}$ ,  $\|\hat{W}^* - W_0\|_{b_n} = o_P(\tau_n n^{-1/4})$ ; (ii)  $P^*(\hat{W}^* \in C(X_n)) \xrightarrow{P} 1$ ;

(iii) For any class of functions  $\Psi$  satisfying condition [CL](#) (see [Appendix A](#)) and for  $\phi_t(x) := E\{\psi_t(Z) \mid X = x\}$ ,

$$\sqrt{n}P_n(\hat{W}^* - \hat{W})t_n\psi_t = \sqrt{n}P_n(a - W_0)\phi_t + o_{P^*}(1) \text{ uniformly over } \mathcal{T}.$$

**Assumption 10.** (i) Rates for the weak learners: for  $b_n \in \{t_n^*, t_n\}$ ,  $\|(\hat{m}_{\hat{W}}^{*[0]} - m_{W_0})b_n\|_\infty = o_{P^*}(d_n)$ ,  $\|(\partial \hat{m}_{\hat{W}}^{*[0]} - \partial m_{W_0})b_n\|_\infty = o_{P^*}(d_n)$ ,  $\|(f_{\hat{W}^*} - f_{W_0})b_n\|_\infty = o_{P^*}(d_n)$ ,  $\|(\partial f_{\hat{W}^*} - \partial f_{W_0})b_n\|_\infty = o_{P^*}(d_n)$ , with  $\frac{d_{n,m}}{\tau_n^B} = o(n^{-1/4})$ .

(ii) Regularity of the weak learner:  $P^*(\hat{m}_{\hat{W}}^{*[0]} \in \mathcal{C}_d) \xrightarrow{P} 1$ ;

(iii) Regularity of the boosting updates: for  $b = 1, \dots, B - 1$ :  $P^*_{\frac{\hat{T}_\varepsilon^*}{f_{W_0}}} \in \mathcal{C}_n \xrightarrow{P} 1$ ;

(iv) Regularity of the sample estimates:  $P^*(\hat{m}_{\hat{W}}^{[B]}(\hat{W}(\cdot)), \frac{\hat{m}_{\hat{W}}^{[B]}(\hat{W}(\cdot))}{\hat{W}_1(\cdot)}, \hat{W} \in \mathcal{C}(\mathcal{X}_n)) \xrightarrow{\Sigma} 1$ ,  $\log N_{[\cdot]}(\delta, \mathcal{C}(\mathcal{X}_n), L_1(P_X)) \leq C\delta^{-v}$ , with  $v \in (0, 1)$ .

Assumption [8](#) has a technical nature and is needed to simplify the proofs. Assumptions [9](#) and [10](#) can be considered the bootstrap equivalents of Assumptions [6](#) and [7](#). In particular, Assumption [9](#) establishes the convergence rates and an expansion for the bootstrap counterpart of the first-step estimator  $\hat{W}^*$ . If the first-step estimator is specified according to the  $L_2$  boosting procedure, Assumption [9](#) can be proved along the arguments of [Appendix B](#) and [C](#). Assumption [10](#) essentially replicates Assumption [7](#) and refers to the second-step estimation in the “bootstrap world”. The only part that is conceptually different is Point (iv) which only imposes a regularity conditions of the sample estimators, so it can be considered as a mild requirement.

The expansion at the basis of our bootstrap test is contained in the following

**Proposition 7.** Under Assumptions [1-10](#), if  $H_0$  holds

$$\sqrt{n}P_n(\hat{\varepsilon}_B^* \hat{t}_n \hat{f}_{\hat{W}^*}(\hat{W}^*)\varphi_t) = \sqrt{n}P_n(\tilde{Y}^* m_{W_0}(W_0))f_{W_0}\varphi_t^\perp - \sqrt{n}P_n(\tilde{a}^* - W_0)\partial m_{W_0}f_{W_0}\varphi_t^\perp + o_{P^*}(1) \\ \text{uniformly over } \mathcal{T},$$

where  $\tilde{Y}^* = \tilde{a}_1^* \cdot \tilde{a}_2^*$  and for any set  $A \subset \{0, 1\}^2$ ,  $P((\tilde{a}_1^*, \tilde{a}_2^*) \in A \mid X) = P^{H_0}((a_1, a_2) \in A \mid X)$ .

The previous proposition provides the validity of the bootstrap test. In particular, consider an arbitrary finite collection  $(t_1, \dots, t_Q) \in \mathcal{T}$ . Then, wpt1 and conditionally on the sample data, the vector  $\sqrt{n}P_n(\hat{\varepsilon}_B^* \hat{t}_n \hat{f}_{\hat{W}^*}(\hat{W}^*)\varphi_{t_q})_{q=1, \dots, Q}$  will converge in distribution to  $G(t_q)_{q=1, \dots, Q}$ , where  $G$  is the Gaussian process defined in [Proposition 6](#). Hence, the critical value simulated by bootstrapping  $S_n^*$  can be

considered as a reasonable approximation of the true critical value of  $S_n^*$ <sup>4</sup>. From a practical point of view, the critical value can be obtained by a Monte-Carlo simulation: (i) simulate  $N^*$  samples from Algorithm 3; (ii) for the  $c$ -th simulated sample, compute the statistic  $S_{n,c}^*$ ; (iii) use the  $1 - \alpha$  quantile of the collection  $\{S_{n,c}^*: c = 1, \dots, N^*\}$  as the critical value for running the test at the  $\alpha$  critical level, so reject  $H_0$  if  $S_n$  is larger than such a quantile. The larger  $N^*$ , the better will be the approximation of the quantile of the bootstrap distribution.

## 7 Simulation and implementation

In this section, we show how to implement the test, and we provide the results of a Monte Carlo experiment to assess its small-sample performances.

### 1. Analysis of the size

We start by analyzing the capacity of the test to control the size under the null  $H_0$ . We assume the same framework as the one presented in Example 2, and for the reader's convenience we report here the main parts of the game-theoretical model. We consider two agents, each denoted with  $s \in \{1, 2\}$ , that must take a binary decision  $a_s \in \{0, 1\}$ . The profit of agent  $s$  writes as

$$\pi_s(a_{-s}, \theta_s) = a_s \cdot [\beta_s x - \alpha_s a_{-s} - \theta_s],$$

where  $x$  is a vector of exogenous covariates observed by both players and the researcher,  $\theta_s$  is the private information of player  $s$ , and  $(\alpha_s, \beta_s)$  are fixed parameters unknown to the researcher.  $(\theta_1, \theta_2) \sim N(0, \Sigma)$ , where  $\Sigma$  is a 2x2 covariance matrix,  $\text{Var}\theta_1 = \text{Var}\theta_2 = 1$ ,  $\text{Cov}(\theta_1, \theta_2) = \rho$ . We require  $\rho \in (-1, \sqrt{2\pi/(2\pi + |\alpha_1|^2)})$ . Define  $(\theta_1^*(x), \theta_2^*(x))$  to be the solution of the system

$$\beta_s x - \alpha_s \Phi((\theta_{-s}^*(x) - \rho \cdot \theta_s^*(x)) / \sqrt{1 - \rho^2}) = \theta_s^*(x), \text{ for } s = 1, 2,$$

where  $\Phi$  denotes the cdf of the standard normal. Along the lines of Example 2, the pure-strategy Bayesian-Nash equilibrium is given by the the profile

$$\delta_s(x, \theta_s) = 1\{\theta_s^*(x) \geq \theta_s\} \text{ for } s = 1, 2.$$

We parameterize the model as follows:  $\beta_1 = (-1, 1), \beta_2 = (1, -1), \alpha_1 = 1, \alpha_2 = 1, x = (x_1, x_2)$ ,  $x_1 \sim x_2, x_1 \perp x_2, x_1 \sim U[-1, 2, 1, 2]$ . The correlation coefficient  $\rho$  is set to different values so that

<sup>4</sup>To make this argument more rigorous, we should prove that the leading term in the expansion of Proposition 7 converges to the Gaussian process  $G$  uniformly over  $T$  and for almost all trajectories  $\{X_1\}_{i \geq 1}$ . This proof is still in progress. However, since  $S_n$  can be expressed as a Riemann Integral, it can also be approximated by a functional of a the finite dimensional process  $\sqrt{n} P_n \varepsilon_B^* \hat{t}_n \hat{f}_{W^*}(\hat{W}^*) \varphi_{t_q}$  for large enough  $Q$ . So, the convergence of the “finite dimensional marginals” provides an argument to conclude that the bootstrap procedure gives a reasonable approximation of the distribution of  $S_n$ .

$\rho \in \{-0.5, 0, 0.5\}$ . Thus, when  $\rho = 0$  there is no correlation among agents' private information, while whenever  $\rho \neq 0$  agents' types are correlated.

For the implementation of the test, in the first step estimation we use the same type of estimator as the second step. So, also for the estimation of the conditional probabilities  $W_0$  we employ a boosting algorithm. We use kernel estimators both for the initial weak learners -needed for the initialization of the  $L_2$  boosting procedure- and for the updating part of Algorithm 2. For the first-step estimation, we set

$$\hat{W}_s^{[0]}(x) = \frac{\hat{T}_s(x)}{\hat{f}(x)}, \hat{T}_s(x) = \frac{1}{n \cdot h^p} \sum_{i=1}^n a_{s,i} \cdot K \cdot \frac{X_i - x}{h}, \hat{f}(x) = \frac{1}{n \cdot h^p} \sum_{i=1}^n K \cdot \frac{X_i - x}{h}$$

for  $s = 1, 2$ . The kernel  $K$  is set to be a second-order Gaussian kernel. Similarly, the kernel used in the Boosting Algorithm 2 is set to be a second-order Gaussian kernel. Both the bandwidth  $h$  above and the bandwidths used in the  $L_2$  boosting procedure of Algorithm 2 are specified according to a Silverman's Rule of Thumb, i.e.  $h = (\hat{\sigma}(X_1), \hat{\sigma}(X_2)) \cdot n^{-1/6}$ . We use a similar logic for the specification of the estimator  $\hat{m}_{\hat{W}^{[B]}}^{[b]}$ . In particular, the weak learner initializing Algorithm 2 is specified as

$$\hat{m}_{\hat{W}^{[B_1]}}^{[0]}(w) := \frac{\hat{T}_{\hat{W}^{[B_1]}}(w)}{\hat{f}_{\hat{W}^{[B_1]}}(w)}, \hat{T}_{\hat{W}^{[B_1]}}(w) = \frac{1}{n \cdot h^d} \sum_{i=1}^n Y_i \cdot K \cdot \frac{\hat{W}_i^{[B_1]} - w}{h},$$

$$\hat{f}_{\hat{W}^{[B_1]}}(w) = \frac{1}{n \cdot h^d} \sum_{i=1}^n K \cdot \frac{\hat{W}_i^{[B_1]} - w}{h},$$

where  $Y_i := a_{1,i} \cdot a_{2,i}$ ,  $K$  is a second-order Gaussian Kernel, and  $h$  is specified according to Silverman's Rule of Thumb, i.e.  $h = (\hat{\sigma}(\hat{W}_1^{[B_1]}), \hat{\sigma}(\hat{W}_2^{[B_1]})) \cdot n^{-1/6}$ . For the updating part of Algorithm 2 we use the same specification for both the bandwidth and the kernel. The number of boosting iterations is set to the same level for both the first and the second step estimation. The iterations employed are displayed in Table 1-5.

For the computation of the ICM statistic  $S_n$ , the weighting function  $\mu$  is defined as a triangular density, so that  $\phi_\mu$  is a sinc kernel.

We compare the *Multinomial Bootstrap* presented in Section 6 to the *Wild-Bootstrap* procedure in ? developed for semiparametric models with generated variables. For this latter bootstrap procedure, we specify the weights  $\{\xi_i\}_{i=1}^n$  to be distributed as iid standard normals, independently from the sample data.

The results of the simulations are reported in Table 1-3. With  $B = 1$  both tests under-reject compared to the nominal size. The test based on the Multinomial Bootstrap behave the best when the number of boosting iterations is set to  $B = 2$ . In general, both the Wild-Bootstrap test and the Multinomial test behave relatively well. In particular, the wild bootstrap has a tendency to under-reject and needs a larger sample to adjust the empirical size to the nominal one. The Multinomial Bootstrap shows a smaller error in the rejection probability for all sample sizes. Intuitively, since the Multinomial Bootstrap is based on a resampling scheme that reflects the null DGP better than the Wild Bootstrap, it also



Table 1: Simulation results for  $B_1 = B_2 = 1$ 

| $n$ |                    | 0.02   | 0.05   | 0.10   |
|-----|--------------------|--------|--------|--------|
| 200 | <b>Wild</b>        | 0.0146 | 0.04   | 0.0962 |
|     | <b>Multinomial</b> | 0.0128 | 0.0328 | 0.0768 |
| 400 | <b>Wild</b>        | 0.0122 | 0.0354 | 0.0896 |
|     | <b>Multinomial</b> | 0.0116 | 0.0318 | 0.0728 |
| 600 | <b>Wild</b>        | 0.0058 | 0.0256 | 0.0732 |
|     | <b>Multinomial</b> | 0.0058 | 0.0226 | 0.0626 |

Simulations based on 5000 Monte-Carlo replications. The test is based on a Warp-speed method for the Monte Carlo simulations of the bootstrap.  $h_0 = (sd(X_1), sd(X_2)) \cdot n^{-1/6}$ ,  $h = (sd(\hat{W}(X))) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

displays a better performance with respect to the Wild scheme.

Generally, when the number of boosting iterations ( $B_1$  and  $B_2$ ) are set to either 0 or 1, the performance of the test is less good<sup>5</sup>. This therefore shows that the boosting is a valid tool to control the size of the test, by controlling the bias of the empirical process at the basis of the statistic  $S_n$ .

## 7.2 Power Analysis

In this section we provide the power analysis of the test. To have a DGP displaying a clear departure from the null hypothesis, we start from Example 3. For the reader's convenience, we report below the main parts of the model. We consider a two-players game and assume that the profit of each agent writes as

$$\Pi_s(x, \eta, a_1, a_2, \theta_s) = a_s \cdot [\psi(x, \eta) - \alpha_s \cdot a_{-s} - \theta_s],$$

with  $\psi(x, \eta) = \phi(x_1) + \eta \cdot \delta(x_2)$ ,  $\theta_1 \sim U[0, 1]$ ,  $\theta_2 \sim U[0, 1]$ ,  $\theta_1 \perp \theta_2$ ,  $(\theta_1, \theta_2) \perp x$ ,  $E\eta = 0$ ,  $\eta \perp x$ ,  $x = (x_1, x_2)$ ,  $x_1 \perp x_2$ . The Pure-Strategy Bayesian-Nash equilibrium of the game is represented by the following functions

$$a_s = \{\psi(x, \eta) + \alpha_s \cdot \mu_{-s}^*(x, \eta) \geq \theta_s\} \text{ for } s = 1, 2, \text{ where } \mu_s^*(x, \eta) = \frac{1 + \alpha_s}{1 - \alpha_1 \cdot \alpha_2} \cdot \psi(x, \eta).$$

$\eta$  is interpreted as a heterogeneity parameter. Let  $(a_1, a_2)$  be generated by the above game, and assume the researcher only observes  $(a_1, a_2, x)$  but not the heterogeneity parameter  $\eta$ . She aims at testing whether the distribution of the data, say  $P_{(a, a)}|_x$ , can be rationalized by a unique Bayesian-Nash equi-

<sup>5</sup>Results are omitted for reason of space



Table 2: Simulation results for  $B_1 = B_2 = 2$ 

| $n$ |             | 0.02   | 0.05   | 0.10   |
|-----|-------------|--------|--------|--------|
| 200 | Wild        | 0.0058 | 0.0206 | 0.0528 |
|     | Multinomial | 0.0184 | 0.0478 | 0.105  |
| 400 | Wild        | 0.0046 | 0.0166 | 0.0464 |
|     | Multinomial | 0.0168 | 0.05   | 0.104  |
| 600 | Wild        | 0.0022 | 0.0104 | 0.0328 |
|     | Multinomial | 0.0128 | 0.043  | 0.0956 |

Simulations based on 5000 Monte-Carlo replications. The test is based on a Warp-speed method for the Monte Carlo simulations of the bootstrap.  $h_0 = (sd(X_1), sd(X_2)) \cdot n^{-1/6}$ ,  $h = (sd(\hat{W}(X))) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

Table 3: Simulation results for  $B_1 = B_2 = 2$ : Comparison accross different correlation coefficients.

| $\rho$ |             | 0.02   | 0.05   | 0.10   |
|--------|-------------|--------|--------|--------|
| 0.5    | Wild        | 0.003  | 0.0104 | 0.0366 |
|        | Multinomial | 0.0118 | 0.0366 | 0.0872 |
| 0      | Wild        | 0.0046 | 0.0166 | 0.0464 |
|        | Multinomial | 0.0168 | 0.05   | 0.104  |
| -0.5   | Wild        | 0.002  | 0.0144 | 0.0426 |
|        | Multinomial | 0.0186 | 0.0558 | 0.1102 |

Simulations based on 5000 Monte-Carlo replications. The test is based on a Warp-speed method for the Monte Carlo simulations of the bootstrap.  $h_0 = (sd(X_1), sd(X_2)) \cdot n^{-1/6}$ ,  $h = (sd(\hat{W}(X))) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

Table 4: Simulation results for  $B_1 = B_2 = 2$ , uniform private information, mutually independent types: Empirical rejection probabilities

| $n$  |                    | 0.02   | 0.05   | 0.10   |
|------|--------------------|--------|--------|--------|
| 500  | <b>Wild</b>        | 0.01   | 0.033  | 0.074  |
|      | <b>Multinomial</b> | 0.0265 | 0.0655 | 0.122  |
| 700  | <b>Wild</b>        | 0.0075 | 0.0335 | 0.0695 |
|      | <b>Multinomial</b> | 0.0195 | 0.049  | 0.1240 |
| 1000 | <b>Wild</b>        | 0.012  | 0.04   | 0.0775 |
|      | <b>Multinomial</b> | 0.026  | 0.0555 | 0.1175 |

Simulations based on 1000 Monte-Carlo replications. The test is based on a Warp-speed method for the Monte Carlo simulations of the bootstrap.  $h_0 = (sd(X_1), sd(X_2)) \cdot n^{-1/6}$ ,  $h = (sd(\hat{W}(X))) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

Table 5: Simulation results for  $B_1 = B_2 = 2$ , uniform private information, mutually independent types: Power Analysis

| $n$  |                    | 0.02   | 0.05   | 0.10   |
|------|--------------------|--------|--------|--------|
| 500  | <b>Wild</b>        | 0.2265 | 0.3755 | 0.5165 |
|      | <b>Multinomial</b> | 0.4145 | 0.552  | 0.685  |
| 700  | <b>Wild</b>        | 0.3595 | 0.532  | 0.6705 |
|      | <b>Multinomial</b> | 0.542  | 0.702  | 0.8165 |
| 1000 | <b>Wild</b>        | 0.528  | 0.699  | 0.824  |
|      | <b>Multinomial</b> | 0.763  | 0.864  | 0.9195 |

Simulations based on 1000 Monte-Carlo replications. The test is based on a Warp-speed method for the Monte Carlo simulations of the bootstrap.  $h_0 = (sd(X_1), sd(X_2)) \cdot n^{-1/6}$ ,  $h = (sd(\hat{W}(X))) \cdot n^{-1/6}$ . The kernels are Gaussian Kernels of order 2.

librium of a game which respects Assumptions [E](#), [D](#), and [M](#). So, the goal is to test whether the condition  $E\{a_1 \cdot a_2 | X\} = E\{a_1 a_2 | W_0(X)\}$  holds in the data. By the discussion provided in [Example 3](#), such an equality does not hold for the DGP considered here, and the departure from the null hypothesis is represented by the difference

$$E\{a_1 \cdot a_2 | x\} - E\{a_1 \cdot a_2 | W_0(x)\} = \frac{(1 + \alpha_1) \cdot (1 + \alpha_2)}{(1 - \alpha_1 \cdot \alpha_2)^2} \cdot E\{\eta^2\} \cdot \delta(x_2)^2 - E\{\delta(x_2)^2\} \cdot \Sigma.$$

Hence, as long as  $\delta(x_2)$  is a non-degenerate function, we are under  $H_1$ . The magnitude of the departure from the null hypothesis is here represented by the variance of the right hand side of the above display: the larger such a variance the greater the departure from  $H_0$ .

We parameterize the model as follows:  $x_1 \sim U[b_1, b_2]$ ,  $x_2 \sim U[b_3, b_4]$ ,  $\phi(x) = c + x$ ,  $\delta(x) = \delta \cdot \eta \cdot x$ ,  $\frac{\gamma}{2}$   
 $(b_1, b_2, b_3, b_4) = (-0.1, 0, 0, 1)$ ,  $\alpha_1 = \alpha_2 = 0.1$ ,  $c = 0.5$ ,  $\gamma = 0.5$ ,  $P(\eta = 0.4) = P(\eta = -0.4) = 0.5$ .

The test is implemented similarly to the previous section. The results are reported in [Table 4](#) and [5](#). Under the null  $H_0$ , the Wild-Bootstrap scheme displays a tendency to under-reject, similarly to what was happening in the previous experiment. The error in the rejection probability seems to be contained for the Multinomial Bootstrap. Under  $H_1$  both tests show a satisfying power, but the capacity of detecting departures from the null hypothesis is more pronounced for the Multinomial Bootstrap than for the Wild Bootstrap. This is an adding feature in support of the Multinomial scheme presented in this work.

## 8 Conclusions

In this paper we provide a test to check if the distribution of the observed data can be rationalized by a unique Pure Strategy Bayesian-Nash Equilibrium of a game with incomplete information. Each player is assumed to take a binary decision, and agents' private types are possibly correlated. We start from a characterization of the null hypothesis in terms of a conditional moment restriction. This involves nonparametric conditional probabilities identified from the data. We propose an Integrated Conditional Moment test. The statistic is based on a nonparametric function of the conditional probabilities that are themselves nonparametrically defined. Hence, we proceed to a two-step estimation: in a first step we estimate these conditional probabilities, while in a second step we estimate the function to be plugged into the statistic. To handle the bias arising from this two steps nonparametric estimation, we use an  $L_2$  boosting algorithm. We show that under the null hypothesis the proposed statistic converges to a functional of a Gaussian process. The asymptotic null distribution depends on unknown features of the data, so it cannot be used to approximate the critical value of the test. We therefore propose a new Multinomial Bootstrap scheme which incorporates all the restrictions  $H_0$  imposes on the distribution of the data. In particular, such a scheme imposes that the data is generated by a distribution that can be rationalized by a unique Bayesian-Nash equilibrium of an incomplete information game. Thanks to the  $L_2$  boosting algorithm, the test can be implemented without undersmoothing. In a Monte Carlo experiment we show that the test has satisfying performances in small samples, both in terms of size

and power:

Several extensions that could be done for this work. We have proposed a Bierens' type of test which, from a technical point of view, can detect a sequence of Pitman alternatives converging to the null at the rate of  $n^{-1/2}$  <sup>6</sup>. It would also be interesting to analyze the properties of a double-smoothing test. This would complement our approach and might be useful from a practical point of view, since it might serve as a tool to give a stronger confirmation of findings when testing the validity of the Bayesian-Nash assumption.

## References

- Aguirregabiria, Victor, & Mira, Pedro. 2007. Sequential Estimation of Dynamic Discrete Games. *Econometrica*, **75**(1), 1–53.
- Ahn, Hyungtaik. 1997. Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors. *Econometric Theory*, **13**(01), 3–31.
- Ahn, Hyungtaik, & Powell, James L. 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, **58**(1-2), 3–29.
- Andrews, Donald W. K. 1995. Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory*, **11**(03), 560–586.
- Aradillas-Lopez, Andres. 2010. Semiparametric estimation of a simultaneous game with incomplete information. *Journal of Econometrics*, **157**(2), 409–431.
- Aradillas-Lopez, Andres. 2012. Pairwise-difference estimation of incomplete information games. *Journal of Econometrics*, **168**(1), 120–140.
- Bajari, Patrick, Hong, Han, Krainer, John, & Nekipelov, Denis. 2010a. Estimating Static Models of Strategic Interactions. *Journal of Business & Economic Statistics*, **28**(4), 469–482.
- Bajari, Patrick, Hong, Han, Krainer, John, & Nekipelov, Denis. 2010b. Estimating Static Models of Strategic Interactions. *Journal of Business & Economic Statistics*, **28**(4), 469–482.
- Berry, Steven. 1992. Estimation of a Model of Entry in the Airline Industry. *Econometrica*, **60**(4), 889–917.
- Bierens, Herman J., & Ploberger, Werner. 1997. Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica*, **65**(5), 1129–1152.

---

<sup>6</sup>Although this has not been shown formally, we expect that our test will share this feature

- Bjorn, Paul A., & Vuong, ang H. 1985 (July). *Econometric Modeling of a Stackelberg Game with an Application to Labor Force Participation*. Tech. rept. 577. California Institute of Technology, Division of the Humanities and Social Sciences.
- Blundell, Richard W., & Powell, James L. 2004. Endogeneity in Semiparametric Binary Response Models. *Review of Economic Studies*, **71**(3), 655–679.
- Bresnahan, Timothy, & Reiss, Peter. 1990. Entry in Monopoly Market. *Review of Economic Studies*, **57**(Feb.), 531–53.
- Bresnahan, Timothy F., & Reiss, Peter C. 1991. Empirical models of discrete games. *Journal of Econometrics*, **48**(1), 57–81.
- Brock, William, & Durlauf, Steven. 2007. Identification of binary choice models with social interactions. *Journal of Econometrics*, **140**(1), 52–75.
- Brock, William A., & Durlauf, Steven N. 2001. Discrete Choice with Social Interactions. *The Review of Economic Studies*, **68**(2), 235–260.
- Ciliberto, Federico, & Tamer, Elie. 2009. Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, **77**(6), 1791–1828.
- Cornillon, Pierre-Andre, Hengartner, N. W., & Matzner Lober, E. 2014. Recursive bias estimation for multivariate regression smoothers. *ESAIM: Probability and Statistics*, **18**, 483–502.
- de Paula, Aureo, & Tang, Xun. 2010 (Apr.). *Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information*. PIER Working Paper Archive. Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Delgado, Miguel A., & Manteiga, Wenceslao Gonzalez. 2001. Significance testing in nonparametric regression based on the bootstrap. *The Annals of Statistics*, **29**(5), 1469–1507.
- Di Marzio, Marco, & Taylor, Charles C. 2008. On boosting kernel regression. *Journal of Statistical Planning and Inference*, **138**(8), 2483–2498.
- Escanciano, Juan Carlos, Jacho-Chávez, David T., & Lewbel, Arthur. 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics*, **178**(P3), 426–443.
- Escanciano, Juan Carlos, Jacho Chavez, David, & Lewbel, Arthur. 2016. Identification and estimation of semiparametric two-step models. *Quantitative Economics*, **7**(2), 561–589.
- Fan, Yanqin, & Li, Qi. 1996. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica*, **64**(4), 865–890.

- Hahn, Jinyong, & Ridder, Geert. 2013. Asymptotic Variance of Semiparametric Estimators With Generated Regressors. *Econometrica*, **81**(1), 315–340.
- Hahn, Jinyong, Liao, Zhipeng, & Ridder, Geert. 2018. Nonparametric Two-Step Sieve M Estimation and Inference. *Econometric Theory*, **34**(6), 1281–1324.
- Jia, Panle. 2008. What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry. *Econometrica*, **76**(6), 1263–1316.
- Kooreman, Peter. 1994. Estimation of econometric models of some discrete games. *Journal of Applied Econometrics*, **9**(3), 255–268.
- Lavergne, Pascal, & Vuong, ang H. 1996. Nonparametric Selection of Regressors: The Nonnested Case. *Econometrica*, **64**(1), 207–219.
- Lavergne, Pascal, Maistre, Samuel, & Patilea, Valentin. 2015. A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*, **9**(1), 643–678.
- Lewbel, Arthur, & Tang, Xun. 2015a. Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics*, **189**(1), 229–244.
- Lewbel, Arthur, & Tang, Xun. 2015b. Identification and estimation of games with incomplete information using excluded regressors. *Journal of Econometrics*, **189**(1), 229–244.
- Li, Qi, & Racine, Jeffrey Scott. 2006. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- Liu, Nianqing, Vuong, ang, & Xu, Haiqing. 2017. Rationalization and identification of binary games with correlated types. *Journal of Econometrics*, **201**(2), 249–268.
- Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2012. Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics*, **40**(2), 1132–1170.
- Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2016. Semiparametric Estimation with Generated Covariates. *Econometric Theory*, **32**(05), 1140–1177.
- Mcpherson, Miller, Smith-Lovin, Lynn, & Cook, James. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, **27**(Jan.), 415.
- Newey, Whitney K. 1994. The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, **62**(6), 1349–1382.
- Newey, Whitney K., Powell, James L., & Vella, Francis. 1999. Nonparametric Estimation of Triangular Simultaneous Equations Models. *Econometrica*, **67**(3), 565–604.

- Pagan, Adrian. 1984. Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, **25**(1), 221–247.
- Park, B. U., Lee, Y. K., & Ha, S. 2009. L2 boosting in kernel regression. *Bernoulli*, **15**(3), 599–613.
- Rothe, Christoph. 2009. Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics*, **153**(1), 51–64.
- Seim, Katja. 2006. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, **37**(3), 619–640.
- Soetevent, Adriaan R., & Kooreman, Peter. 2007. A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics*, **22**(3), 599–624.
- Stinchcombe, Maxwell B., & White, Halbert. 1998. Consistent Specification Testing With Nuisance Parameters Present Only Under The Alternative. *Econometric Theory*, **14**(03), 295–325.
- Sweeting, Andrew. 2009. The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *RAND Journal of Economics*, **40**(4), 710–742.
- Tang, Xun. 2010. Estimating simultaneous games with incomplete information under median restrictions. *Economics Letters*, **108**(3), 273–276.
- van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge University Press.
- van der Vaart, A. W., & Wellner, Jon. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.
- Vanhems, Anne, & Keilegom, Ingrid Van. 2019. Estimation of a Semiparametric Transformation Model in the Presence of Endogeneity. *Econometric Theory*, **35**(1), 73–110.

## A Asymptotic Expansion

Before proving the main results, we introduce a class of functions that will be often used in the technical proofs.

**Condition CL.**  $\Psi := \{x \mapsto \psi_t(x) : t \in \mathcal{T}\}$  is a collection of uniformly bounded Lipschitz functions in  $t$ , i.e.  $\|\psi_{t_1} - \psi_{t_2}\|_\infty \leq C \|t_1 - t_2\|$  for all  $t_1, t_2 \in \mathcal{T}$ . We define  $\phi_t(W(X)) := E\{\psi_t(X) | W(X)\}$ .

We start with a technical lemma that will be needed in the derivation of the Brahadur representation of the empirical process at the basis of  $S_n$ .

**Lemma 8. (Stochastic Equicontinuity Results)** Let Assumption 1-7 hold, and let  $\Psi$  and  $\phi_t$  be as in Condition [CL](#). Then, for  $b = 0, \dots, B$  uniformly in  $t \in \mathbb{T}$

- (i)  $\sqrt{n}(\mathbf{P}_n - \mathbf{P}) \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n \psi_t = o_P(1);$
- (ii)  $\sqrt{n}(\mathbf{P}_n - \mathbf{P}) \hat{T}_{\varepsilon_b}(W_0) t_n \psi_t = o_P(1);$
- (iii)  $\sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n \psi_t = \sqrt{n} \mathbf{P} \varepsilon_b \phi_t + \sqrt{n} \mathbf{P} (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[b]}(\hat{W})) t_n \phi_t + h^r \sqrt{n} \mathbf{P}_n \varepsilon^{[b]} t_n \phi_{n,t}^{(1)} + o_P(1);$
- (iv)  $\sqrt{n} \mathbf{P} \hat{T}_{\varepsilon_b}(W_0) t_n \psi_t = \sqrt{n} \mathbf{P}_n \varepsilon_b f_{W_0} + \sqrt{n} \mathbf{P} (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[b]}(\hat{W})) t_n \phi_{n,t} f_{W_0} + h^r \sqrt{n} \mathbf{P}_n \varepsilon^{[b]} t_n \phi_{n,t}^{(1)} f_{W_0} + o_P(1);$
- (v)  $\sqrt{n}(\mathbf{P}_n - \mathbf{P}) (\hat{m}_{\hat{W}}^{[b]}(W_0) - m_{W_0}(W_0)) t_n \psi_t = o_P(1)$ , and the same result holds by replacing  $t_n$  with  $\hat{t}_n$ ,

where  $\phi_{n,t}^{(1)}$  is a function satisfying Condition [CL](#).

*Proof.* For Point (i), by Lemma [24](#), Assumption 4, 5, 6, and 7, we obtain

$$\sup_{t \in \mathbb{T}} \sqrt{n} \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n \psi_t = o_P(1).$$

By the above display, Assumption 7(iii), and Lemma [17](#), we can apply Lemma [19](#) to obtain Point (i).

Point (ii) can be proved along the same lines as Point (i).

For Point (iii), by Assumption 4, Lemma [24](#), and the Law of Iterated Expectations, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} \sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n \psi_t &= \sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n^{W_0} \psi_t + \sqrt{n} \mathbf{P} \frac{\hat{T}_{\varepsilon_b}}{f_{W_0}}(W_0) t_n^{W_0} (t_n - 1) \psi_t \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} \int K(v) \phi_t(\hat{W}_i + vh) t_n^{W_0}(\hat{W}_i + vh) dv + o_P(1) = \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} \int K(v) \phi(\hat{W}_i + vh) dv + o_P(1), \end{aligned} \quad (6)$$

where in the third equality we have used Assumptions 4 and 6 and Markov's inequality to drop the trimming  $t_n^{W_0}$ . Now, by the rates in Lemma [24](#), Lemma [21](#), and Assumption 6,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} \int K(v) \phi_t(\hat{W}_i + vh) dv &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} \int K(v) \phi(\hat{W}_i + vh) dv + o_P(1) = \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} \int K(v) \phi_t(W_{0,i} + vh) dv + \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_{i,b} \hat{t}_{n,i} (\hat{W}_i - W_{0,i}) \int K(v) \partial \phi_t(W_{0,i} + vh) dv + o_P(1) \end{aligned} \quad (7)$$

uniformly in  $t \in \mathbb{T}$ , where in the second equality we have used a Mean-Value expansion of  $\phi(\hat{W}_i + vh)$  around  $W_0 + vh$  and the rates for  $\hat{W}$  in Assumption 6. The convergence rates in Lemma [24](#) and



Assumption 6 imply that uniformly over  $\mathcal{T}$ ,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{i,b} t_{n,i} (\hat{W}_i - W_{0,i}) \int K(v) \partial \phi_t(W_{0,i} + vh) dv = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i t_{n,i} (\hat{W}_i - W_{0,i}) \int K(v) \partial \phi_t(W_{0,i} + vh) dv + o_P(1) = o_P(1), \quad (8)$$

where the second equality follows from the expansion in Assumption 6 and since  $E\{\varepsilon | X\} = 0$ . Finally, by a usual  $r$ -th order Taylor expansion and Assumption 4, uniformly over  $\mathcal{T}$ ,

$$\sqrt{n} P_n \varepsilon \phi_t + \sqrt{n} P_n ((m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[b]}(\hat{W})) t_n \phi_t + h^r \sqrt{n} P_n \varepsilon^{[b]} t_n \phi_{n,t}^{(1)} + o_P(1)). \quad (9)$$

Putting together Eq. 9, 8, 7 and 6 yields Point (ii).

Point (iv) can be proved along the same lines as Point (iii), so the proof is omitted.

For Point (v), by using the recursive structure in Eq. 3,

$$\sqrt{n} (P_n - P) (\hat{m}_{\hat{W}}^{[b]}(W_0) - m_{W_0}(W_0)) t_n \psi_t = \sqrt{n} (P_n - P) (\hat{m}_{\hat{W}}^{[0]}(W_0) - m_{W_0}(W_0)) t_n \psi_t + \sum_{s=0}^{b-1} \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}}{f_{\hat{W}}^{[0]}}(W_0) t_n \psi_t$$

Assumption 7(ii) and Lemma 17 combined with Lemma 19 ensure that the first term on the RHS of the previous display is  $o_P(1)$  uniformly over  $\mathcal{T}$ . For a generic addendum of the second term on the RHS, by the rates in Lemma 24 and Assumption 7,

$$\sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}}{f_{\hat{W}}^{[0]}}(W_0) t_n \psi_t = \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}}{f_{W_0}}(W_0) t_n \psi_t + o_P(1)$$

uniformly over  $\mathcal{T}$ . Using Point (i), Lemma 24, and Assumption 5-6, the leading term of the above expression is  $o_P(1)$  uniformly in  $t$ . By this and the two previous display, we conclude for Point (v).  $\square$

**Lemma 9. (First expansion)** Under Assumption 1-7, uniformly over  $\mathcal{T}$ ,

$$\sqrt{n} P_n \hat{\varepsilon}_n \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}) \phi_t = \sqrt{n} P_n \varepsilon f_{W_0} \phi_t^\perp - \sqrt{n} P_n (a - W_0) \partial m_{W_0} f_{W_0} \phi_t^\perp - h^r \sqrt{n} P_n \varepsilon^{[B-1]} t_n \phi_{n,t}^{(1)} f_{W_0} + o_P(1).$$

*Proof.* By Lemma 24, Assumption 7, and Lemma 21 we can replace the trimming  $\hat{t}_n$  with  $t_n$ , so that uniformly over  $\mathcal{T}$ ,

$$\sqrt{n} P_n \hat{\varepsilon}_n \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}) \phi_t = \sqrt{n} P_n \hat{\varepsilon}_n t_n f_{\hat{W}}(\hat{W}) \hat{\phi} + o_P(1) = \sqrt{n} P_n \varepsilon \hat{f}_{\hat{W}}(\hat{W}) t_n \phi_t + \sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n f_{\hat{W}}(\hat{W}) \phi_t + o_P(1). \quad (10)$$

For the first term on the RHS of the previous display, by Assumption 6 and 7, and a Mean-Value expansion of  $\hat{f}_{\hat{W}}(\hat{W})$  around  $W_0$

$$\begin{aligned} \sqrt{n} P_n \varepsilon \hat{f}_{\hat{W}}(\hat{W}) \varphi_t &= \sqrt{n} P_n \varepsilon \hat{f}_{\hat{W}}(W_0) \varphi_t + \\ &\sqrt{n} P_n \varepsilon \partial f_{W_0}(W_0) (\hat{W} - W_0) \varphi_t + o_P(1) \text{ uniformly over } \mathcal{T}. \end{aligned}$$

Using Assumption 7 together with Lemma 19 delivers

$$\sqrt{n} P_n \varepsilon \hat{f}_{\hat{W}}(W_0) \varphi_t = \sqrt{n} P_n \varepsilon f_{W_0}(W_0) \varphi_t + o_P(1)$$

uniformly over  $\mathcal{T}$ . By the expansion in Assumption 6 and since  $E\{\varepsilon | X\} = 0$  under  $H_0$ ,

$$\sqrt{n} P_n \varepsilon \partial f_{W_0}(W_0) (\hat{W} - W_0) \varphi_t = o_P(1) \text{ uniformly over } \mathcal{T}.$$

We now turn to the second term on the RHS of Eq. 10. Using the recursive structure of  $m_{\hat{W}}^{[B]}$ , a Mean-Value expansion of  $\hat{T}_{\hat{\varepsilon}_{B-1}}(\hat{W})$  around  $W_0$ , and the rates in Lemma 24 and Assumption 6,

$$\begin{aligned} &\sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \hat{f}_{\hat{W}}(\hat{W}) \varphi_t = \\ &\sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-1]}(\hat{W})) t_n \hat{f}_{\hat{W}}(\hat{W}) \varphi_t - \sqrt{n} P_n \hat{T}_{\hat{\varepsilon}_{B-1}}(W_0) t_n \varphi_t + o_P(1) = \\ &\sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-1]}(\hat{W})) t_n f_{W_0}(W_0) \varphi_t - \sqrt{n} P_n \hat{T}_{\hat{\varepsilon}_{B-1}}(W_0) t_n \varphi_t + o_P(1) \end{aligned}$$

uniformly over  $\mathcal{T}$ , where in the last equality we have used the convergence rate of  $\hat{m}_{\hat{W}}^{[B-1]}$  in Lemma 24 and Assumption 7. Using Lemma 8 for the second term on the RHS of the previous display,

$$\begin{aligned} \sqrt{n} P_n \hat{T}_{\hat{\varepsilon}_{B-1}}(W_0) t_n \varphi_t &= \sqrt{n} P_n \varepsilon t_n f_{W_0} + \sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-1]}(\hat{W})) t_n t_n f_{W_0} + \\ &h^r \sqrt{n} P_n \varepsilon^{[B-1]} t_n \phi_{n,t}^{(1)} f_{W_0} + o_P(1), \end{aligned}$$

uniformly over  $\mathcal{T}$ . Replacing the previous displays into Eq. 10 delivers

$$\begin{aligned} \sqrt{n} P_n \varepsilon \hat{f}_{\hat{W}}(\hat{W}) \varphi_t &= \sqrt{n} P_n \varepsilon f_{W_0}(W_0) \varphi_t + \\ &\sqrt{n} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-1]}(\hat{W})) t_n f_{W_0}(W_0) \varphi_t - h^r \sqrt{n} P_n \varepsilon^{[B-1]} t_n \phi_{n,t}^{(1)} f_{W_0} + o_P(1) \end{aligned}$$

uniformly over  $\mathcal{T}$ . By the recursive structure of  $\hat{m}_{\hat{W}}^{[B-1]}$ , the rates in Lemma 24 and Assumption 7,

$$\begin{aligned} &\sqrt{n} P_n (\hat{m}_{\hat{W}}^{[B-1]}(\hat{W}) - m_{W_0}(W_0)) t_n f_{W_0}(W_0) \varphi_t = \\ &\sqrt{n} P_n (\hat{m}_{\hat{W}}^{[0]}(\hat{W}) - m_{W_0}(W_0)) t_n f_{W_0}(W_0) \varphi_t + \sum_{s=1}^{B-1} \sqrt{n} P_n \hat{T}_{\hat{\varepsilon}_s}(W_0) t_n \varphi_t + o_P(1) = \\ &\sqrt{n} P_n (\hat{m}_{\hat{W}}^{[0]}(W_0) - m_{W_0}(W_0)) t_n f_{W_0}(W_0) \varphi_t + \sum_{s=0}^{B-1} \sqrt{n} P_n \hat{T}_{\hat{\varepsilon}_s}(W_0) t_n f_{W_0} \varphi_t + \\ &\sqrt{n} P_n \partial m_{W_0} f_{W_0} \varphi_t (\hat{W} - W_0) + o_P(1) \text{ uniformly over } \mathcal{T}, \end{aligned}$$

where in the last equality we have used a first order expansion of  $m_{\hat{W}}^{[0]}(\hat{W})$  around  $W_0$ , Assumption 5, 6 and 7. Now, notice that  $\sqrt{n} \mathbf{P}_n \mathcal{F}_{\hat{\xi}}(W_0)_{t_{W_0} \phi_t}$  is a centered process by the Law of Iterated Expectations, so using Lemma 8, the expansion in Assumption 6, and the two previous displays we conclude.  $\square$

**Lemma 10. (Negligibility of the boosting iterations)** Under Assumptions 1-7, uniformly in  $t \in \mathcal{T}$ ,

$$h^r \sqrt{n} \mathbf{P}_n \mathcal{E}_{\hat{\xi}}^{[B-1]} t_n \phi_{n,t}^{(1)} f_{W_0} = o_P(1).$$

*Proof.* We first show the following recursive structure:

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n \mathcal{E}_{\hat{\xi}^{B-s}} t_n \tilde{\beta}_{n,t} &= -h^{r(s+1)} \sqrt{n} \mathbf{P}_n \mathcal{E}_{\hat{\xi}^{B-(s+1)}} t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1) \\ &\text{for } s = 1, \dots, B-1 \end{aligned} \quad (11)$$

uniformly over  $\mathcal{T}$ , for a weighting function  $\beta_{n,t}$  satisfying the CL condition. To this end, by the recursive structure of  $m_{\hat{W}}^{[b]}$ , the rates in Assumption 6 and 7, and Lemma 24,

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n \mathcal{E}_{\hat{\xi}^{B-s}} t_n \tilde{\beta}_{n,t} &= h^{rs} \sqrt{n} \mathbf{P}_n \mathcal{E} t_n \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-s]}(\hat{W})) t_n \tilde{\beta}_{n,t} = \\ &h^{rs} \sqrt{n} \mathbf{P}_n \mathcal{E} t_n \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-s]}(W_0)) t_n \tilde{\beta}_{n,t} \\ &\quad - h^{rs} \sqrt{n} \mathbf{P}_n \partial m_{W_0} \tilde{\beta}_{n,t} t_n (\hat{W} - W_0) + o_P(1) \end{aligned} \quad (12)$$

uniformly over  $\mathcal{T}$ . Since  $\tilde{\beta}_{n,t}$  satisfies condition CL, by an application of Lemma 19,

$$h^{rs} \sqrt{n} \mathbf{P}_n \mathcal{E} t_n \tilde{\beta}_{n,t} = o_P(1) \text{ uniformly over } \mathcal{T}. \quad (13)$$

Similarly, by the expansion in Assumption 6,

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n \partial m_{W_0} \tilde{\beta}_{n,t} t_n (\hat{W} - W_0) &= h^{rs} \sqrt{n} \mathbf{P}_n \partial m_{W_0} \tilde{\beta}_{n,t} t_n (a - W_0) + o_P(1) \\ &\text{uniformly over } \mathcal{T}. \end{aligned} \quad (14)$$

For the second term on the RHS of Eq. 12, by the recursive structure of  $m_{\hat{W}}^{[b]}$ , the rates in Assumption 6 and 7, and Lemma 24,

$$\begin{aligned} h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-s]}(W_0) - m_{W_0}(W_0)) t_n \tilde{\beta}_{n,t} &= h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-(s+1)]}(W_0) - m_{W_0}(W_0)) t_n \tilde{\beta}_{n,t} + \\ &h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\xi}^{B-(s+1)}}}{f_{W_0}}(W_0) t_n \tilde{\beta}_{n,t} + o_P(1) \text{ uniformly over } \mathcal{T}. \end{aligned}$$

Using Lemma 8, the second term on the RHS of the previous display can be approximated as

$$\begin{aligned}
h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{T}_{\hat{\varepsilon}_{B-(s+1)}}}{f_{W_0}}(W_0) t_n \tilde{\beta}_{n,t} &= h^{rs} \sqrt{n} \mathbf{P}_n \varepsilon \tilde{\beta}_{n,t} + h^{rs} \sqrt{n} \mathbf{P}_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-(s+1)]}(\hat{W})) t_n \tilde{\beta}_{n,t} + \\
&\quad + h^{r(s+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(s+1)} t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1) = \\
h^{rs} \sqrt{n} \mathbf{P}_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[B-(s+1)]}(W_0)) t_n \tilde{\beta}_{n,t} &+ h^{r(s+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(s+1)} t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1) \\
&\text{unif. over } \mathbb{T},
\end{aligned}$$

where  $\tilde{\beta}_{n,t}(w) := E\{\tilde{\beta}_{n,t}(X) | W(X) = w\}$ , and in the second equality we have used the same arguments as in Eq. 12, 13, and 14. Putting together the previous two displays yields

$$\begin{aligned}
&h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-s]}(W_0) - m_{W_0}(W_0)) t_n \tilde{\beta}_{n,t} = \\
&h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-(s+1)]}(W_0) - m_{W_0}(W_0)) t_n \tilde{\beta}_{n,t}^\perp + h^{r(s+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(s+1)} t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1)
\end{aligned}$$

uniformly over  $\mathbb{T}$ , where  $\tilde{\beta}_{n,t}^\perp(X) = \tilde{\beta}_{n,t}(X) - \tilde{\beta}_{n,t}(W_0(X)) = \tilde{\beta}_{n,t}(X) - E\{\tilde{\beta}_{n,t}(X) | W(X)\}$ . By the Assumption 5, we can replace the trimming  $t_n$  with  $t_n^{W_0}$  in the first term on the RHS of the above expression, so that

$$\begin{aligned}
&h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-(s+1)]}(W_0) - m_{W_0}(W_0)) t_n \tilde{\beta}_{n,t}^\perp = \\
&h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B-(s+1)]}(W_0) - m_{W_0}(W_0)) t_n^{W_0} \tilde{\beta}_{n,t}^\perp = o_P(1) \text{ unif. over } \mathbb{T},
\end{aligned}$$

where the second equality follows from noticing that the process on the RHS is centered (by definition of  $\tilde{\beta}_{n,t}^\perp$ ) and applying Lemma 8. By the previous two displays, Eq. 12, 13, and 14, we obtain that Eq. 11 holds. By the recursive structure in Eq. 11 and a simple induction we obtain

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1} t_n \tilde{\beta}_{n,t} = -h^{rB} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_0 t_n \tilde{\beta}_{n,t}^{(1)} + o_P(1)$$

uniformly over  $\mathbb{T}$ . Conclude by the above expression and the rates in Assumption 5 and 7.  $\square$

**Observation:** The above lemma shows that, under Assumptions 1-6, the boosting iterations do not have an impact on the Bahadur expansion of the empirical process  $\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{f}_{\hat{W}}(\hat{W}) \varphi$  at the basis of the statistic  $S_n$ . This results and the asymptotic distribution of  $S_n$  are reported in the following Corollary.

**Corollary 11. (Proof of Proposition 6)** Under Assumption 1-7,

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B \hat{f}_{\hat{W}}(\hat{W}) \varphi_t = \sqrt{n} \mathbf{P}_n \varepsilon f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n (a - W_0) \partial m_{W_0} f_{W_0} \varphi_t^\perp + o_P(1)$$

uniformly over  $\mathbb{T}$ . Accordingly,

$$S_n = \int_0^1 \frac{1}{\sqrt{n}} P_n \varepsilon_B \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}) \varphi_t^2 \mu(dt) + \int_0^1 G \mu(dt),$$

where  $G$  is a Gaussian stochastic process taking values in  $A^\infty(G)$ , with  $G := (a, y, x) \mapsto (y - m_{W_0}(W_0(x))) \cdot f_{W_0}(W_0(x)) \cdot \varphi_t^\perp(x) - (a - W_0(x)) \partial m_{W_0}(W_0(x)) f_{W_0}(W_0(x)) \varphi_t^\perp(x) : t \in \mathbb{T}$  and defined by the collection of covariances  $P \varepsilon^2 \varphi_{t_1}^\perp \varphi_{t_2}^\perp : t_1, t_2 \in \mathbb{T}$ .

*Proof.* The first result is an immediate consequence of Lemma 9 and 10. For the second result, we can proceed as in Example 19.7 in van der Vaart (1998) by the compactness of  $\mathbb{T}$  and the boundedness of the random variables involved, to obtain that  $N(\delta, G, \|\cdot\|_\infty) \leq N(C\delta, \mathbb{T}, \|\cdot\|)$ , with  $\int_0^1 \sqrt{\log N(\delta, \mathbb{T}, \|\cdot\|)} d\delta < \infty$ . By Theorem 19.14 in van der Vaart (1998),  $\sqrt{n} P_n \varepsilon \varphi_t^\perp \rightarrow G$ . Hence, conclude by Theorem 18.11(i) in van der Vaart (1998).  $\square$

## B Bootstrap Expansion

We first reformulate the bootstrap DGP in a way that is relatively easy to handle from a mathematical point of view. Define the following DGP:

$$\begin{aligned} a_1^* &= 1' \hat{W}_1(X_i) \geq u_{1,i}^*, \\ a_2^* &= 1' a_{1,i}^* = 1' \cdot 1' \frac{\hat{m}_2^{[B]}(\hat{W}_1(X_i))}{\hat{W}_1(X_i)} \geq u_{2,i}^* + 1' a_{1,i}^* = 0' \cdot 1' \frac{\hat{W}_2(X_i) - \hat{m}_2^{[B]}(\hat{W}_1(X_i))}{1 - \hat{W}_1(X_i)} \geq u_{2,i}^{**}, \\ Y_i^* &= a_{1,i}^* \cdot a_{2,i}^* \text{ and } u^* := (u_{1,i}^*, u_{2,i}^*, u_{2,i}^{**}) \sim \text{iid } U[0, 1]. \end{aligned} \quad (15)$$

The bootstrap DGP defined in Eq. 15 generates the same probability measure for  $(a_1^*, a_2^*)$  as Algorithm 3. In this sense the two DGPs are equivalent. Working with the DGP in Eq. 15 is more convenient from a mathematical viewpoint, so we choose to build our proofs on it. Define  $P^*$  the probability measure that considers only  $u^*$  as random and the sample data as fixed. Recall that  $P$  denotes the probability measure of the sample. Let  $P := P^* \otimes P$  be the probability measure that considers both  $u^*$  and the sample data as random.

**Remark.** Notice that given a real-valued random variable  $g(u^*, X)$  and an arbitrary set  $A \subset \mathbb{R}$ , the probability  $P^*(g(X, u^*) \in A)$  is random.

**Lemma 12.** Let  $\psi_t, \phi_t$ , and  $\phi_{n,t}^{(1)}$  be as in Lemma 8 and let Assumptions 1-10 hold. Then for  $b = 1, \dots, B$  and uniformly over  $\mathbb{T}$ ,

- (i)  $\sqrt{n}(P_n - P) \frac{\hat{T}_{\varepsilon_b}^*(W_0)}{f_{W_0}} t_n \psi_t = o_{P^*}(1);$
- (ii)  $\sqrt{n}(P_n - P) \hat{T}_{\varepsilon_b}^*(W_0) t_n \psi_t = o_{P^*}(1);$

$$\begin{aligned}
(iii) \quad & \sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n \psi_t = \sqrt{n} P_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \phi_t(W) + \\
& \sqrt{n} P_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \phi_t(W) + o_P(1); \\
(iv) \quad & \sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n \psi_t = \sqrt{n} P_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \phi_t(W) + \\
& + \sqrt{n} P_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \phi_t(W) + o_P(1); \\
(v) \quad & \sqrt{n} (P_n - P) (\hat{m}_{\hat{W}^*}^{[B]}(W_0) - \hat{m}_{\hat{W}}^{[B]}(W_0)) t_n \psi_t = o_P(1)
\end{aligned}$$

*Proof.* For Point (i), by Lemma 26

$$\sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n \psi_t = o_P(1).$$

By the previous display and Assumption 10, we can apply Lemma 20 to obtain the desired result.

The proof of Point (ii) proceeds in the same way as the proof of Point (i).

For Point (iii), from the rates in Lemma 26, Assumption 4, and the Law of Iterated Expectations,

$$\begin{aligned}
\sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n \psi_t &= \sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n^{W_0} \psi_t + \sqrt{n} P_{\frac{\hat{T}_{\varepsilon_b}^*}{f_{W_0}}(W_0)} t_n^{W_0} (t_n - 1) \psi_t = \\
&= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(\hat{W}_i^* + vh) t_n^{W_0} (\hat{W}_i^* + vh) dv + o_P(1) = \\
&= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(\hat{W}_i^* + vh) dv + o_P(1) \text{ unif. over } \mathbb{T},
\end{aligned}$$

where in the second equality we have used Assumption 6 and Markov's inequality to drop the trimming  $t_n^{W_0}$ . Using the rates in Lemma 26 and Assumption 4-9,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(\hat{W}_i^* + vh) dv &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(\hat{W}_i^* + vh) dv + o_P(1) = \\
&= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(W_{0,i} + vh) dv + \\
&+ \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \partial \phi_t(W_{0,i} + vh) (\hat{W}_i^* - W_{0,i}) dv + o_P(1)
\end{aligned} \tag{16}$$

uniformly over  $\mathbb{T}$ , where in the second equality we have used a Mean-Value expansion of  $\phi(\hat{W}_i^* + vh)$  around  $\phi_t(W_{0,i} + vh)$  and the rates in Assumption 9. For the second term on the RHS of the previous display, Lemma 26, Lemma 24, and Assumption 9 ensure that

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \partial \phi_t(\hat{W}_i^* + vh) dv &= \\
\sqrt{n} P_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_{n,i} (\hat{W}_i^* - W_{0,i}) \int K(v) \partial \phi_t(W_{0,i} + vh) dv &= o_P(1) \\
&\text{unif. over } \mathbb{T},
\end{aligned}$$

where in the last equality we have used Assumption 9 and Lemma 25. For the first term on the RHS of

Eq. 16, by the usual  $r$ -th order Taylor expansion

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{b,i}^* \int K(v) \phi_t(W_{0,i} + vh) dv = \\ & \sqrt{n} P_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \phi_t(W_{0,i}) + \sqrt{n} P_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \phi_t(W_{0,i}) + \\ & h^r \sqrt{n} P_n \hat{\varepsilon}_{b,i}^* \phi_{n,t}^{(1)} + o_{P^*}(1) \text{ unif. over } \mathbb{T}. \end{aligned}$$

Conclude for Point (iii) by the previous four displays.

The proof of Point (iv) proceeds along the same lines as the proof of Point (iii).

For Point (iv), by the recursive structure of  $\hat{m}_{\hat{W}}^{[b]}$  and  $\hat{m}_{\hat{W}^*}^{*[b]}$

$$\begin{aligned} & \sqrt{n} (P_n - P) (\hat{m}_{\hat{W}^*}^{*[b]}(W_0) - \hat{m}_{\hat{W}}^{[b]}(W_0)) t_n \psi_t = \sqrt{n} (P_n - P) (\hat{m}_{\hat{W}^*}^{*[0]}(W_0) - \hat{m}_{\hat{W}}^{[0]}(W_0)) t_n \psi_t + \\ & \sum_{s=0}^{B-1} \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}^*}{f_{\hat{W}^*}}(W_0) t_n \psi_t + \sum_{s=0}^{B-1} \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}}{f_{\hat{W}}} (W_0) t_n \psi_t. \end{aligned}$$

The first term on the RHS of the previous display is  $o_{P^*}(1)$  uniformly over  $\mathbb{T}$ , by Assumption 10 and the compactness of  $\mathbb{T}$ . Proceeding as in the proof of Point (v) of Lemma 8 yields that the third term on the RHS of the previous display is  $o_{P^*}(1)$  uniformly over  $\mathbb{T}$ . By the rates in Lemma 24, Lemma 26, and Assumption 7-10,

$$\begin{aligned} & \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}^*}{f_{\hat{W}^*}}(W_0) t_n \psi_t = \sqrt{n} (P_n - P) \frac{\hat{T}_{\varepsilon_s}^*}{f_{W_0}}(W_0) t_n \psi_t + o_{P^*}(1) = \\ & o_{P^*}(1) \text{ uniformly over } \mathbb{T}, \end{aligned}$$

where in the second equality we have used Point (i) of the present Lemma. Conclude by the previous two displays.  $\square$

**Lemma 13.** *Let Assumptions 1-10 hold, and  $\Psi$  be a class of functions that respects condition CL. Then, uniformly over  $\mathbb{T}$ ,*

$$\sqrt{n} P_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \psi_t = \sqrt{n} P_n(\tilde{Y}^* - m_{W_0}(W_0)) t_n \psi_t + o_{P^*}(1),$$

where  $\tilde{Y}^* := \mathbf{1}' W_{0,1}(X) \geq u_1^* \cdot \mathbf{1}' \frac{m_{W_0}(W_0)}{W_{0,1}}(X) \geq u_2^*$ .

*Proof.* By proceeding as in Point (i) of Lemma 25,

$$\sup_{t \in \mathbb{T}} \dots (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \psi_t - (\tilde{Y}^* - m_{W_0}(W_0)) t_n \psi_t \dots_{L_2(P)} = o_{P^*}(1).$$

By the previous display, Assumption 10, and Lemma 18, an immediate application of Lemma 20 yields the desired result.  $\square$

**Lemma 14. (A First Bootstrap Expansion)** Let Assumptions 1-10 hold. Then, uniformly over  $\mathbb{T}$

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B^* \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t = \sqrt{n} \mathbf{P}_n(\tilde{Y}^* - m_{W_0}(W_0)) f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n(\tilde{a}^* - W) \partial m_{W_0} f_{W_0} \varphi_t^\perp + h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \phi_{n,t} + o_{P^*}(1),$$

where  $\tilde{Y}^* := \mathbf{1}' W_{0,1}(X) \geq u_1^* \cdot \mathbf{1}' \frac{m_{W_0}(W_0)}{W_{0,1}}(X) \geq u_2^*$ ,  $\tilde{a}^* := (\tilde{a}_1^*, \tilde{a}_2^*)$ ,  $\tilde{a}_1^* := \mathbf{1}' W_{0,1}(X) \geq u_1^*$ ,  $\tilde{a}_2^* := \mathbf{1}' \tilde{a}_1^* = \mathbf{1}' \cdot \mathbf{1}' \frac{m_{W_0}(W_0(X))}{W_{0,1}(X)} \geq u_2^* + \mathbf{1}' \tilde{a}_1^* = 0 \cdot \mathbf{1}' \frac{W_{0,2}(X) - m_{W_0}(W_0(X))}{1 - W_{0,1}(X)} \geq u_2^{**}$ .

*Proof.* Using the rates in Lemma 26, Lemma 21, and Assumption 9-10,

$$\begin{aligned} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B^* \hat{t}_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t &= \sqrt{n} \mathbf{P}_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t + \\ &\sqrt{n} \mathbf{P}_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t + o_{P^*}(1) \text{ unif. over } \mathbb{T}. \end{aligned} \quad (17)$$

Using a Mean-Value Expansion of  $\hat{f}_{\hat{W}}(\hat{W}^*)$  around  $W_0$  Assumption 9-10, and Lemma 25,

$$\begin{aligned} \sqrt{n} \mathbf{P}_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t &= \\ \sqrt{n} \mathbf{P}_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \hat{f}_{\hat{W}}(W_0) + \partial f_{W_0}(W_0) \cdot (\hat{W}^* - W_0) \varphi_t + o_{P^*}(1) &= \\ \sqrt{n} \mathbf{P}_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n f_{W_0}(W_0) \varphi_t + o_{P^*}(1) \text{ unif. over } \mathbb{T}. \end{aligned} \quad (18)$$

By Lemma 24, Lemma 26, and Assumption 9-10, uniformly over  $\mathbb{T}$ ,

$$\begin{aligned} \sqrt{n} \mathbf{P}_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t &= \\ \sqrt{n} \mathbf{P}_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B-1]}(\hat{W}^*)) t_n f_{W_0} \varphi_t - \sqrt{n} \mathbf{P}_n \frac{T_{\hat{W}^*}^{[B-1]}}{f_{\hat{W}^*}}(\hat{W}^*) t_n f_{W_0} \varphi_t + o_{P^*}(1). \end{aligned}$$

Using Lemma 26, Assumption 9-10, and Lemma 12 on the second term RHS of the previous display

$$\begin{aligned} \sqrt{n} \mathbf{P}_n \frac{T_{\hat{W}^*}^{[B-1]}}{f_{\hat{W}^*}}(\hat{W}^*) t_n f_{W_0} \varphi_t &= \sqrt{n} \mathbf{P}_n \hat{T}_{\hat{\varepsilon}_{B-1}^*}^*(W_0) t_n f_{W_0} \varphi_t + o_{P^*}(1) = \\ &\sqrt{n} \mathbf{P} \hat{T}_{\hat{\varepsilon}_{B-1}^*}^*(W_0) t_n f_{W_0} \varphi_t + o_{P^*}(1) \text{ unif. over } \mathbb{T}. \end{aligned}$$

The second term on the RHS of the previous display can be approximated according to Lemma 12. By such approximation and the previous two displays, uniformly over  $\mathbb{T}$

$$\begin{aligned} \sqrt{n} \mathbf{P}_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W}^*)) t_n \hat{f}_{\hat{W}}(\hat{W}^*) \varphi_t &= -\sqrt{n} \mathbf{P}_n(Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \partial f_{W_0} + \\ &\sqrt{n} \mathbf{P}_n(\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{[B-1]}(\hat{W}^*)) t_n f_{W_0} \varphi_t - h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \phi_{n,t}^{(1)} f_{W_0} + o_{P^*}(1) \end{aligned} \quad (19)$$



Using the rates in Lemma 24, Lemma 26, and Assumption 6-7-9-10,

$$\begin{aligned} & - \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-1]}(\hat{W}^*)) t_{\hat{W}_0} f_{W_0} \varphi_t^\perp = \\ & \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(W_0) - \hat{m}_{\hat{W}^*}^{*[B-1]}(W_0)) t_n f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n \partial m_{W_0} f_{W_0} \varphi_t^\perp (\hat{W}^* - W_0) + o_P(1) \\ & \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(W_0) - \hat{m}_{\hat{W}^*}^{*[B-1]}(W_0)) t_n^{W_0} f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n \partial m_{W_0} f_{W_0} \varphi_t^\perp (\hat{W}^* - W_0) + o_P(1) \end{aligned}$$

uniformly over  $\mathcal{T}$ , where in the second equality we have replaced the trimming  $t_n$  with  $t_n^{W_0}$  by Assumption 4, Lemma 24, and Lemma 26. Notice that the first term on the RHS of the previous display is a centered process, by definition of  $\varphi_t^\perp$ . Hence, using Lemma 12 such term will be  $o_P(1)$  uniformly over  $\mathcal{T}$ . Using this result and Assumption 9, from the previous display we obtain

$$\begin{aligned} & - \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-1]}(\hat{W}^*)) t_{\hat{W}_0} f_{W_0} \varphi_t^\perp = \\ & - \sqrt{n} \mathbf{P}_n \partial m_{W_0} f_{W_0} \varphi_t^\perp (\hat{W}^* - W_0) + o_P(1) \text{ uniformly over } \mathcal{T}. \end{aligned} \quad (20)$$

Plugging the above display, Eq. 19, and 18 into Eq. 17 yields

$$\begin{aligned} & \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B^* \hat{t}_{\hat{W}^*}(\hat{W}^*) \varphi_t = \sqrt{n} \mathbf{P}_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n f_{W_0} \varphi_t^\perp + \\ & - \sqrt{n} \mathbf{P}_n \partial m_{W_0} f_{W_0} \varphi_t^\perp (\hat{W}^* - W_0) - h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B^* t_n \phi_{n,t}^{(1)} f_{W_0} + o_P(1) \\ & \text{uniformly over } \mathcal{T}. \end{aligned}$$

Conclude by using Assumption 9-10 and Lemma 13 on the first term RHS of the previous display.  $\square$

**Lemma 15. (Negligibility of the Reminder Term in Bootstrap Expansion)** Under Assumptions 1-10,

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \phi_{n,t} = o_P(1) \text{ uniformly over } \mathcal{T}.$$

*Proof.* We first obtain the following recursive structure:

$$\begin{aligned} & h^{rs} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-s}^* t_n \beta_{n,t} = -h^{r(s+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(s+1)}^* t_n \beta_{n,t}^{(1)} + o_P(1) \\ & \text{uniformly over } \mathcal{T}, \text{ for } s = 1, \dots, B-1, \end{aligned} \quad (21)$$

for a weight function  $\beta_{n,t}$  satisfying condition CL. Using Lemma 25, Assumption 10, and the recursive structure of  $\hat{m}_{\hat{W}^*}^{*[B]}$ ,

$$\begin{aligned} & h^{rs} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-s}^* t_n \beta_{n,t} = h^{rs} \sqrt{n} \mathbf{P}_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \beta_{n,t} + \\ & h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-s]}(\hat{W}^*)) t_n \beta_{n,t} + o_P(1) = \\ & h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-(s+1)]}(\hat{W}^*)) t_n \beta_{n,t} - h^{rs} \sqrt{n} \mathbf{P}_n \frac{\hat{\varepsilon}_{B-(s+1)}^*}{\hat{f}_{\hat{W}^*}}(\hat{W}^*) t_n \beta_{n,t} + o_P(1) \end{aligned}$$

uniformly over  $\mathcal{T}$ . By the rates in Lemma 26 and Assumptions 9-10,

$$h^{rs} \sqrt{n} \mathbf{P}_n \frac{T_{\hat{\varepsilon}_{B-(s+1)}^*}^*}{f_{\hat{W}^*}}(\hat{W}^*) t_n \beta_{n,t} = h^{rs} \sqrt{n} \mathbf{P}_n \frac{T_{\hat{\varepsilon}_{B-(s+1)}^*}^*}{f_{W_0}}(W_0) t_n \beta_{n,t} + o_{P^*}(1) \\ \text{uniformly over } \mathcal{T}.$$

Applying Lemma 12 to the leading term of the above display and using the two previous displays yields

$$h^{rs} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-s}^* t_n \beta_{n,t} = h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}^*}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-(s+1)]}(\hat{W}^*)) t_n \beta_{n,t}^\perp + \\ - h^{r(s+1)} \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-(s+1)}^* t_n \beta_{n,t}^{(1)} + h^r \sqrt{n} \mathbf{P}_n (Y^* - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W})) t_n \tilde{\beta}_{n,t} + o_{P^*}(1) \\ \text{uniformly over } \mathcal{T},$$

where  $\tilde{\beta}_{n,t}(\cdot) = E\{\beta_{n,t}(X) \mid W_0(X) = \cdot\}$  and  $\beta_{n,t}^\perp(X) := \beta_{n,t}(X) - \tilde{\beta}_{n,t}(W_0(X))$ . By Lemma 25 the second term on the RHS of the previous expression is  $o_{P^*}(1)$  uniformly in  $t \in \mathcal{T}$ . Using a reasoning similar to Eq. 20 delivers

$$h^{rs} \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}^*}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[B-(s+1)]}(\hat{W}^*)) t_n \beta_{n,t}^\perp = \\ - h^{rs} \sqrt{n} \mathbf{P}_n \partial m_{W_0} f_{W_0} \varphi_t^*(\tilde{a}^* W) + o_{P^*}(1) = \\ o_{P^*}(1) \text{ uniformly over } \mathcal{T}.$$

The previous two displays prove the recursive structure in Eq. 21. By such recursive structure and a simple induction,

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_{B-1}^* t_n \beta_{n,t} = -h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_0^* t_n \beta_{n,t}^{(1)} + o_{P^*}(1) \text{ uniformly over } \mathcal{T}.$$

The first term on the RHS of the previous display can be decomposed as

$$h^r \sqrt{n} \mathbf{P}_n \hat{\varepsilon}_0^* t_n \beta_{n,t}^{(1)} = \\ h^r \sqrt{n} \mathbf{P}_n (Y^* - \hat{m}_{\hat{W}^*}^{[B]}(\hat{W})) t_n \beta_{n,t}^{(1)} + h^r \sqrt{n} \mathbf{P}_n (\hat{m}_{\hat{W}^*}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[0]}(\hat{W}^*)) t_n \beta_{n,t}^{(1)}.$$

By Lemma 25 the first term on the RHS of the previous display is  $o_{P^*}(1)$  uniformly over  $\mathcal{T}$ . The rates in Assumptions 6-10 and Lemma 24 ensure that the second term is also negligible. So, we conclude.  $\square$

**Corollary 16.** (Proof of Proposition 7) Under Assumptions 1-10,

$$\sqrt{n} \mathbf{P}_n \hat{\varepsilon}_B^* \hat{t}_n \hat{f}_{\hat{W}^*}(\hat{W}^*) \varphi_t = \sqrt{n} \mathbf{P}_n (\tilde{Y}^* - m_{W_0}(W_0)) f_{W_0} \varphi_t^\perp - \sqrt{n} \mathbf{P}_n (\tilde{a}^* - W) \partial m_{W_0} f_{W_0} \varphi_t^\perp + \\ \text{uniformly over } \mathcal{T},$$

where  $\tilde{Y}^* := 1' W_{0,1}(X) \geq u_1^*$ ,  $\tilde{a}^* := (\tilde{a}_1^*, \tilde{a}_2^*)$ ,  $\tilde{a}_1^* := 1' W_{0,1}(X) \geq u_1^*$ ,

$$\tilde{a}_2^* := 1' \tilde{a}_1^* = 1' \cdot 1' \frac{m_{W_0}(W_0(X))}{W_{0,1}(X)} \geq u_2^* + 1' \tilde{a}_1^* = 0' \cdot 1' \frac{W_{0,2}(X) - m_{W_0}(W_0(X))}{1 - W_{0,1}(X)} \geq u_2^{**}.$$

*Proof.* The result immediately follows from Lemma 14 and Lemma 15.  $\square$

## C Auxiliary Lemmas

**Lemma 17. (Entropy bounds)** Denote with  $Z$  a random variable taking values in  $\mathcal{Z}$ , with  $P_Z$  its probability measure, and with  $\Psi$  a class of functions satisfying condition CL. Let  $\mathbf{F}_n$  be a sequence of classes of uniformly bounded functions defined over  $\mathcal{Z}$ . It holds that,

$$(i) N(\delta, \Psi \mathbf{F}_n, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{F}_n, \|\cdot\|_\infty) \cdot N(C\delta, \mathbf{T}, \|\cdot\|);$$

$$(ii) \text{ If } \log N(\delta, \mathbf{F}_n \Psi, \|\cdot\|_\infty) \leq C_1 \delta^{-v}, \text{ then } \log N(\delta, \mathbf{F}_n \Psi, \|\cdot\|_\infty) \leq C_2 \delta^{-v}.$$

*Proof.* For Point (i), given two classes  $\mathbf{F}$  and  $\mathbf{A}$  of uniformly bounded functions defined on an arbitrary set  $\mathcal{Y}$ , it is immediate to see that  $N(\delta, \mathbf{F} \mathbf{A}, \|\cdot\|_{\infty, \mathcal{Y}}) \leq N(C\delta, \mathbf{F}, \|\cdot\|_{\infty, \mathcal{Y}}) \cdot N(C\delta, \mathbf{A}, \|\cdot\|_{\infty, \mathcal{Y}})$  for a finite constant  $C$ . Using the Lipschitz property of the class  $\Psi$  and by proceeding as in Example 19.7 of [van der Vaart \(1998\)](#),  $N(\delta, \Psi, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{T}, \|\cdot\|)$ . The previous three inequalities imply that  $N(\delta, \Psi \mathbf{F}_n, \|\cdot\|_\infty) \leq N(C\delta, \mathbf{F}_n, \|\cdot\|_\infty) \cdot N(\delta, \mathbf{T}, \|\cdot\|)$ . Point (ii) follows from this last inequality, the compactness of  $\mathbf{T}$  and Theorem 2.7.1 in [van der Vaart & Wellner \(1996\)](#) (see also the Observation in Section 5).  $\square$

**Lemma 18.** Let  $P_Z$  be the probability measure of a variable  $Z$ ,  $P_U$  be the probability measure for  $U \sim U[0, 1]$ , and  $\mathbf{P} := P_Z \otimes P_U$ . Define the class of functions  $\mathbf{F} := \{(u, z) \mapsto 1\{m(x) \geq u\} : m \in \mathbf{M}\}$ , where  $\mathbf{M}$  is a collection of functions with range in  $(\eta, 1 - \eta)$  with  $\eta \in (0, 1)$ . Then,

$$N_{[\cdot]}(\sqrt{\delta}, \mathbf{F}, L_2(\mathbf{P})) \leq N_{[\cdot]}(\delta, \mathbf{M}, L_1(P_Z)).$$

*Proof.* Fix  $\delta > 0$ , and consider a collection of brackets covering  $\mathbf{M}$ , say  $\mathbf{A} := \{[l_j, s_j] : j = 1, \dots, N_{[\cdot]}(\delta, \mathbf{M}, L_1(P_Z))\}$  with  $\|s_j - l_j\|_{L_1(P_Z)} < \delta$ . Given the range of the elements in  $\mathbf{M}$ , for each bracket  $[l_j, s_j]$  we can take  $l_j$  and  $s_j$  with range in  $(0, 1)$  without loss of generality. Define  $\tilde{l}_j(u, z) := 1\{u \leq l_j(z)\}$ ,  $\tilde{s}_j(u, z) := 1\{u \leq s_j(z)\}$ , and the collection of brackets  $\tilde{\mathbf{A}} := \{[\tilde{l}_j, \tilde{s}_j] : j = 1, \dots, N_{[\cdot]}(\delta, \mathbf{M}, L_1(P_Z))\}$ . Pick an arbitrary element  $f \in \mathbf{F}$ , say  $f(u, z) = 1\{u \leq m(z)\}$  with  $m \in \mathbf{M}$ . By definition of  $\mathbf{A}$ ,  $m \in [l_j, s_j]$  for some bracket in  $\mathbf{A}$ , so that  $l_j(z) \leq m(z) \leq s_j(z)$  and  $1\{u \leq l_j(z)\} \leq 1\{u \leq m(z)\} \leq 1\{u \leq s_j(z)\}$  for all  $(u, z)$  in the support of  $\mathbf{P} = P_U \otimes P_Z$ . By the arbitrariness of  $f$ , the collection  $\tilde{\mathbf{A}}$  covers the class  $\mathbf{F}$ . For the size of each bracket in  $\tilde{\mathbf{A}}$ , by definition of  $\mathbf{P}$ ,  $(\tilde{l}_j, \tilde{s}_j)$ , and Tonelli-Fubini's Theorem,

$$\|\tilde{s}_j - \tilde{l}_j\|_{L_2(\mathbf{P})}^2 = \int \int 1\{l_j(z) \leq u \leq s_j(z)\} dP_U dP_Z = \int |s_j(z) - l_j(z)| dP_Z = \|s_j - l_j\|_{L_1(P_Z)} < \delta.$$

Conclude by the previous display.  $\square$

**Lemma 19.** Assume that  $Z$ ,  $P_Z$ ,  $F_n$ , and  $\Psi$  are as in Lemma 17. Let  $f_0$  be a fixed function defined over  $Z$ , and  $\hat{f}$  be a random function over  $Z$ , where the randomness is considered wrt the probability  $P_Z$ . Define  $G_n := \sqrt{n}(P_n - P_Z)$ . If

- (i)  $\|\hat{f} - f_0\|_{L_2(P_Z)} = o_P(1)$ ,
- (ii)  $P(\hat{f} \in F_n) \rightarrow 1$ , with  $\log N_{[\cdot]}(\delta, F_n, L_2(P_Z)) \leq C\delta^{-v}$  and  $v \in (0, 2)$ ,

then

$$G_n(\hat{f} - f_0)\psi_t = o_P(1) \text{ uniformly over } T.$$

*Proof.* Define  $\tilde{F}_n := (F_n - f_0)\Psi$ . Since the entropy of  $F_n - f_0$  is equal to that of  $F_n$ , Lemma 17 implies that  $\log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$ . Fix  $\delta > 0$ . By Assumptions (i) and (ii) of the present Lemma, for an arbitrary  $\eta > 0$ ,

$$\limsup_{n \rightarrow \infty} P \cdot \sup_{t \in T} G_n(\hat{f} - f_0)\psi_t > \eta \leq \limsup_{n \rightarrow \infty} P \cdot \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta,$$

where  $\tilde{F}_n(\delta) := \{f \in \tilde{F}_n : \|f\|_{L_2(P_Z)} < \delta\}$ . The RHS of the previous display can be upperbounded according to the maximal inequality in Lemma 19.34 of [van der Vaart \(1998\)](#). Since  $\log N_{[\cdot]}(\delta, \tilde{F}_n(\delta), L_2(P_Z)) \leq \log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$  with  $v \in (0, 2)$ , we can choose a small enough  $\delta$  to make such upper-bound arbitrarily small. By the arbitrariness of  $\eta$ , we conclude.  $\square$

**Lemma 20.** Assume that  $Z$ ,  $P_Z$ ,  $F_n$ ,  $\Psi$ , and  $f_0$  are as in Lemma 19. Let  $\hat{f}$  be a random function over  $Z$  where the randomness is considered wrt a probability  $P = P_Z \otimes P_U^*$ , with  $P_U^*$  being a probability measure. Define  $G_n := \sqrt{n}(P_n - P)$ . If

- (i)  $\|\hat{f} - f_0\|_{L_2(P)} = o_{P_U^*}(1)$
- (ii)  $P_U^*(\hat{f} \in F_n) \xrightarrow{P} 1$ , with  $\log N_{[\cdot]}(\delta, F_n, L_2(P_Z)) \leq C\delta^{-v}$  and  $v \in (0, 2)$

then

$$G_n(\hat{f} - f_0)\psi = o_{P_U^*}(1) \text{ uniformly over } T.$$

*Proof.* By the same reasoning as in Lemma 19,  $\log N_{[\cdot]}(\delta, \tilde{F}_n, L_2(P_Z)) \leq C\delta^{-v}$ . Fix  $\delta > 0$ . Assumptions (i) and (ii) of the present Lemma ensure that for an arbitrary  $\eta > 0$ ,

$$\begin{aligned} P_U^* \cdot \sup_{t \in T} G_n(\hat{f} - f_0)\psi_t > \eta &\leq P_U^* \cdot \sup_{t \in T} G_n(\hat{f} - f_0)\psi_t > \eta \leq \\ &P_U^* \cdot \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta + P_U^* \cdot \sup_{t \in T} \|(\hat{f} - f_0)\psi_t\|_{L(P)} > \delta + P_U^* \cdot \sup_{\hat{f} \notin F_n} G_n \hat{f} > \eta \\ &P_U^* \cdot \sup_{f \in \tilde{F}_n(\delta)} G_n f > \eta + o_P(1). \end{aligned}$$

For the first term on the RHS of the above display, using Markov's inequality

$$P_Z \cdot P_U^* \sup_{f \in \tilde{F}_n(\delta)} G_n f \cdot > \eta \cdot \sum \leq \frac{1}{\eta} E_{P_Z} \cdot P_U^* \sup_{f \in \tilde{F}_n(\delta)} G_n f \cdot > \eta \cdot \sum, = \frac{1}{\eta} P \sup_{f \in \tilde{F}_n(\delta)} G_n f \cdot > \eta \cdot \sum,$$

where the last inequality follows from the product structure of  $P$ , i.e.  $P = P_Z \otimes P_U^*$ , and Tonelli-Fubini's Theorem. With the same arguments as in the proof of Lemma 19, we can show that by choosing  $\delta$  small enough the RHS of the previous display can be made arbitrarily small. So, by the arbitrariness of  $\eta$  and the previous two displays we conclude.  $\square$

**Lemma 21. (Trimming)** Under Assumption 4 and 6(i),

$$(i) \sqrt{n} P_n | \hat{t}_n - t_n | = o_P(1)$$

(ii) If  $\sup_{t \in T} \| \hat{g}_t \|_\infty = O_P(1)$  and  $\sup_{t \in T} \| \hat{g}_t \|_\infty = O_P(1)$ , then  $\sqrt{n} P_n \hat{g}(\hat{t}_n - t_n) = o_P(1)$  uniformly over  $T$ .

*Proof.* Let  $B_n := \{ \| \hat{f}^{[0]} - f \|_\infty \leq C d_{n,f} \}$  and fix  $\delta > 0$  arbitrarily small. Assumption 6(i) ensures that, by choosing  $C$  large enough,  $P(B_n) > 1 - \delta$  for any large  $n$ . By definition of  $\hat{t}_n$  we can write

$$\hat{t}_n(w) = 1 \cdot f(w) \geq \tau_n \cdot 1 - \frac{\hat{f}^{[0]}(w) - f(w)}{\tau_n} \cdot \sum.$$

If the event  $B_n$  holds and  $n$  is large enough so that  $C d_{n,f} / \tau_n \leq 1/2$ ,

$$1 - C \frac{d_{n,f}}{\tau_n} \leq 1 - \frac{\hat{f}^{[0]}(w) - f(w)}{\tau_n} \leq 1 + C \frac{d_{n,f}}{\tau_n} \leq \frac{3}{2}$$

for all  $x \in X$ . For such  $n$  and when the event  $B_n$  holds, by the two previous displays

$$f(w) \geq \frac{3}{2} \tau_n \Rightarrow \hat{t}_n(w) = t_n(w) = 1.$$

Hence, when  $B_n$  holds and  $C d_{n,f} / \tau_n \leq 1/2$ , we have  $| \hat{t}_n - t_n | (w) \leq 1 \cdot f(w) \leq \frac{3}{2} \tau_n$ . Using this, Markov's inequality, and Assumption 4,

$$P \cdot \sqrt{n} P_n | \hat{t}_n - t_n | > \delta \leq P \cdot \sqrt{n} P_n | \hat{t}_n - t_n | > \delta \cap B_n \cdot \sum + \delta \leq \delta^{-1} \sqrt{n} \cdot P \cdot f(W) \leq \frac{3}{2} \tau_n \cdot \sum + \delta = o(1) + \delta.$$

By the arbitrariness of  $\delta$  we conclude for point (i). Point (ii) follows immediately from point (i) and after noticing that  $\hat{t}_n - t_n \stackrel{P}{\rightarrow} 0$   $\hat{t}_n \stackrel{P}{\rightarrow} t_n$   $\hat{t}_n \stackrel{P}{\rightarrow} t_n$   $\hat{t}_n \stackrel{P}{\rightarrow} t_n$ .  $\square$

**Lemma 22.** Let  $\{Z_{i,n}\}_{i=1}^n$  be a triangular array of real-valued random variables such that  $|Z_{n,i}| \leq C$  for all  $n$  and  $i$ , and let  $L$  be a kernel function that is Lipschitz continuous. If  $\frac{\log n}{nh^d} = o(1)$ , then

$$\frac{1}{nh^d} \sum_{i=1}^n Z_{n,i} L\left(\frac{W_i - W}{h}\right) - h^{-d} \mathbb{E}\left[Z_{n,i} L\left(\frac{W - W_0}{h}\right)\right] = O_p\left(\frac{\log n}{nh^d}\right).$$

*Proof.* The proof is the same as Theorem 1.4 in [Li & Racine \(2006\)](#) (pages 36-40).  $\square$

**Lemma 23.** Under Assumption 1, 2, and 6,

- (i)  $\sqrt{n} \mathbb{P}_{n \varepsilon} t_n \partial K\left(\frac{w - W}{h}\right) = O_p(1)$  uniformly in  $W$ ;
- (ii) If  $\hat{g}_n$  is a sequence of function defined on  $\mathbf{X}$  such that  $P(\hat{g}_n \in \mathcal{C}(\mathbf{X})) \rightarrow 1$  and  $\|\hat{g}_n\|_\infty = o_p(1)$ , then  $\sqrt{n} \mathbb{G}_n \hat{g}_n t_n \partial K\left(\frac{w - W_0}{h}\right) = o_p(1)$  uniformly over  $W$ .
- (iii) If  $\hat{g}_n$  is a sequence of function defined on  $\mathbf{W}$  such that  $P(\hat{g}_n \in \mathcal{C}(\mathbf{W})) \rightarrow 1$  and  $\|\hat{g}_n\|_\infty = o_p(1)$ , then  $\sqrt{n} \mathbb{G}_n \hat{g}_n(W_0) t_n \partial K\left(\frac{w - W_0}{h}\right) = o_p(1)$  uniformly over  $W$ .

*Proof.* For Point (i), fix  $\delta > 0$  and consider the  $\delta$  covers  $\mathbf{A}_1 := \{w_j : j = 1, \dots, N(\delta, \mathbf{W}, \|\cdot\|)\}$ , and  $\mathbf{A}_2 := \{W^{(j)} : j = 1, \dots, N(\delta, \mathcal{C}(\mathbf{X}_n), \|\cdot\|_\infty)\}$ . Then, for arbitrary  $w \in \mathbf{W}$  and  $W \in \mathcal{C}(\mathbf{X}_n)$ , we must have  $\|w - w_1\| < \delta$  and  $\|W - W^{(1)}\|_{\mathbf{X}_n} < \delta$  for some  $w_1 \in \mathbf{A}_1$  and  $W_1 \in \mathbf{A}_2$ . Since by Assumption 3  $|\partial K(u_1) - \partial K(u_2)| \leq C\|u_1 - u_2\|$  for any  $u_1, u_2$ ,

$$\left| \partial K\left(\frac{w - W(x)}{h}\right) - \partial K\left(\frac{w_1 - W^{(1)}(x)}{h}\right) \right| \leq C\|w - w_1\| + C\|W - W^{(1)}\|_{\mathbf{X}_n} \leq C\delta.$$

The previous display and the arbitrariness of  $\delta$  ensure that the class of functions  $\mathcal{U} := \left\{ \varepsilon t_n(x) \partial K\left(\frac{w - W(x)}{h}\right) : w \in \mathbf{W} \text{ and } W \in \mathcal{C}(\mathbf{X}_n) \right\}$  satisfies the following entropy bound  $N(\delta, \mathcal{U}, \|\cdot\|_\infty) \leq \#\mathbf{A}_1 \cdot \#\mathbf{A}_2 = N(\delta, \mathbf{W}, \|\cdot\|) \cdot N(\delta, \mathcal{C}(\mathbf{X}_n), \|\cdot\|_\infty)$ . By definition of  $\mathcal{C}(\mathbf{X}_n)$ , Theorem 2.7.1 in [van der Vaart & Wellner \(1996\)](#), and Example 19.7 in [van der Vaart \(1998\)](#), the previous inequality implies that

$$\log N_{[\cdot]}(\delta, \mathcal{U}, \|\cdot\|_\infty) \leq \log N(\delta/2, \mathcal{U}, \|\cdot\|_\infty) \leq C\delta^{-v} \text{ with } v \in (0, 2).$$

Assumption 6, the previous two displays, and Lemma 19 ensure that

$$\sqrt{n} \mathbb{P}_{n \varepsilon} t_n \partial K\left(\frac{w - W}{h}\right) = \sqrt{n} \mathbb{P}_{n \varepsilon} \partial K\left(\frac{w - W_0}{h}\right) + o_p(1) \text{ uniformly over } W.$$

Since  $\partial K$  is uniformly bounded and  $h = o(1)$ , using again Lemma 19 ensures that the RHS of the previous display is negligible uniformly in  $w$ .

The proofs of Point (ii) and Point (iii) proceed in a similar way as above by using Assumption 7.  $\square$

**Lemma 24.** Under Assumption 5 and 6(i), uniformly over  $\mathbf{W}$ ,

$$\begin{aligned} (i) \hat{T}_{\varepsilon_b}(w) &= O_P \left( \frac{n^{-1/2}}{h^{d+1}\tau_n^b} + \frac{P_n|\hat{t}_n - t_n|}{h^d\tau_n^b} + \frac{\|(\hat{W} - W_0)t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty}{\tau_n^b} \right) \Sigma; \\ (ii) \partial \hat{T}_{\varepsilon_b}(w) &= O_P \left( \frac{n^{-1/2}}{h^{d+2}\tau_n^b} + \frac{P_n|\hat{t}_n - t_n|}{h^{d+1}\tau_n^b} + \frac{\|(\hat{W} - W_0)t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty}{\tau_n^b} \right) \Sigma \\ (iii) \|(m_{\hat{W}}^{[b]} - m_{W_0})t_n\|_\infty &= O_P \left( \frac{n^{-1/2}}{h^{d+1}\tau_n^b} + \frac{P_n|\hat{t}_n - t_n|}{h^d\tau_n^b} + \frac{\|(\hat{W} - W_0)t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty}{\tau_n^b} \right) \Sigma \end{aligned}$$

where point (iii) holds true also by replacing  $\hat{t}_n$  with  $t_n$ .

*Proof.* By definition of  $\hat{T}_{\varepsilon_b}$

$$\begin{aligned} \partial \hat{T}_{\varepsilon_b}(w) &= \frac{1}{h^{d+1}} P_n \varepsilon t_n \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma + \frac{1}{h^{d+1}} P_n (m_{W_0}(W_0) - m_{\hat{W}}^{[0]}(\hat{W})) t_n \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma + \\ &\quad \frac{1}{h^{d+1}} P_n \varepsilon (\hat{t}_n - t_n) \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma. \end{aligned} \quad (22)$$

For the first term on the RHS of the above display, by Lemma 23,

$$\frac{1}{h^{d+2}} P_n \varepsilon t_n \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma = O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \Sigma \text{ uniformly over } \mathbf{W}. \quad (23)$$

For the third term on the RHS of Eq. 22, since  $\hat{t}_n - t_n = (\hat{t}_n + t_n)(\hat{t}_n - t_n)$ , the rates in Assumption 7(i) ensure that

$$\frac{\sqrt{n}}{h^{d+1}} P_n \varepsilon (\hat{t}_n - t_n) \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma = O_P \left( \frac{P_n |\hat{t}_n - t_n|}{h^{d+1}} \right) \Sigma \text{ uniformly in } w. \quad (24)$$

For the second term on the RHS of Eq. 22, by Assumption 6-7(i) and a first order Taylor expansion of  $m_{\hat{W}}^{[0]}(\hat{W})$  around  $W_0$ ,

$$\begin{aligned} \frac{1}{h^{d+1}} P_n (m_{\hat{W}}^{[0]}(\hat{W}) - m_{W_0}(W_0)) t_n \partial K \left( \frac{w - \hat{W}}{h} \right) \Sigma &= \frac{1}{h^{d+1}} P_n (m_{\hat{W}}^{[0]}(W_0) - m_{W_0}(W_0)) t_n \partial K \left( \frac{w - W_0}{h} \right) \Sigma + \\ &\quad \frac{1}{h^{d+1}} P_n \partial m_{W_0}(W_0) t_n (\hat{W} - W_0) \partial K \left( \frac{w - W_0}{h} \right) \Sigma + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \Sigma \text{ uniformly over } \mathbf{W}. \end{aligned} \quad (25)$$

Lemma 23 and Assumption 6-7 imply that

$$\begin{aligned} \frac{1}{h^{d+1}} P_n (m_{\hat{W}}^{[0]}(W_0) - m_{W_0}(W_0)) t_n \partial K \left( \frac{w - W_0}{h} \right) \Sigma &= \Sigma \\ \frac{1}{h^{d+1}} P_n \partial m_{W_0}(W_0) t_n (\hat{W} - W_0) \partial K \left( \frac{w - W_0}{h} \right) \Sigma &+ O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \Sigma \\ &\text{uniformly over } \mathbf{W}, \end{aligned} \quad (26)$$

By Assumption 2 and Tonelli Fubini's Theorem,

$$\int \left( \hat{m}_{\hat{W}}^{[0]}(W_0(x)) - m_{W_0}(W_0(x)) \right) \cdot t_n(x) \cdot \int \frac{1}{h^{d+1}} \partial K \left( \frac{w - \hat{W}_0}{h} \right) f_{W_0}(\tilde{w}) f_{X|W_0}(x|\tilde{w}) d\tilde{w} dx = O_P \left( \frac{n^{-1/2}}{h^{d+1}} + \|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty \right) \text{ uniformly over } W, \quad (27)$$

where in the second equality we have used an Integration by Parts and the usual change of variable. Putting together Eq. 26 and 27 yields

$$\frac{1}{h^{d+1}} P_n \left( \hat{m}_{\hat{W}}^{[0]}(W_0) - m_{W_0}(W_0) \right) t_n \partial K \left( \frac{w - \hat{W}_0}{h} \right) = O_P \left( \frac{n^{-1/2}}{h^{d+2}} + \|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty \right) \text{ uniformly over } W. \quad (28)$$

By proceeding similarly as for Eq. 28,

$$\frac{1}{h^{d+1}} P_n \partial m_{W_0}(W_0) (\hat{W} - W_0) \partial K \left( \frac{w - \hat{W}_0}{h} \right) = O_P \left( \frac{n^{-1/2}}{h^{d+2}} + \|(\hat{W} - W_0)t_n\|_\infty \right) \text{ uniformly over } W. \quad (29)$$

Plugging Eq. 29, 28, 25, 24, and 23 into 22,

$$\partial \hat{T}_{\varepsilon_0}(w) = O_P \left( \frac{n^{-1/2}}{h^{d+2}} + \frac{P_n \|\hat{t}_n - t_n\|_\infty}{h^{d+1}} + \|(\hat{W} - W_0)t_n\|_\infty + \|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty \right) \text{ uniformly over } W. \quad (30)$$

By a reasoning similar to Eq. 30,

$$\hat{T}_{\varepsilon_0}(w) = O_P \left( \frac{n^{-1/2}}{h^{d+1}} + \frac{P_n \|\hat{t}_n - t_n\|_\infty}{h^d} + \|(\hat{W} - W_0)t_n\|_\infty + \|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty \right) \text{ uniformly over } W. \quad (31)$$

Using the above rate and the recursive structure of  $\hat{m}_{\hat{W}}^{[1]}$

$$\|(\hat{m}_{\hat{W}}^{[1]} - m_{W_0})t_n\|_\infty = \|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n + \frac{\hat{T}_{\varepsilon_0}}{f_{\hat{W}}} t_n\|_\infty = O_P \left( \frac{n^{-1/2}}{h^{d+1}\tau_n} + \frac{P_n \|\hat{t}_n - t_n\|_\infty}{h^d \tau_n} + \frac{\|(\hat{W} - W_0)t_n\|_\infty}{\tau_n} + \frac{\|(\hat{m}_{\hat{W}}^{[0]} - m_{W_0})t_n\|_\infty}{\tau_n} \right) \text{ uniformly over } W. \quad (32)$$

We now obtain the convergence rate for  $\partial \hat{T}_{\hat{\varepsilon}_1}$  and  $\hat{T}_{\hat{\varepsilon}_1}$  By definition of  $\hat{T}_{\hat{\varepsilon}_1}$

$$\partial \hat{T}_{\hat{\varepsilon}_1}(w) = \frac{1}{h^{d+1}} P_n \hat{\varepsilon}_1(\hat{t}_n - t_n) \partial K \left( \frac{w - \hat{W}}{h} \right) + \frac{1}{h^{d+1}} P_n \varepsilon_n t_n \partial K \left( \frac{w - \hat{W}}{h} \right) + \frac{1}{h^{d+1}} P_n (m_{W_0}(W_0) - \hat{m}_{\hat{W}}^{[0]}(\hat{W})) t_n \partial K \left( \frac{w - \hat{W}}{h} \right) - \frac{1}{h^{d+1}} P_n \frac{\hat{T}_{\hat{\varepsilon}_0}}{f_{\hat{W}}}(\hat{W}) t_n \partial K \left( \frac{w - \hat{W}}{h} \right). \quad (33)$$



Using the rates in Eq. 30, 31, Assumption 6-7, and Lemma 23,

$$\begin{aligned} \frac{1}{h^{d+1}} \mathbb{P}_n \frac{\hat{T}_{\varepsilon_0}}{f_{\hat{W}}}(\hat{W}) t_n \partial K \cdot \frac{w-\hat{W}}{h} &= \frac{1}{h^{d+1}} \mathbb{P}_n \frac{\hat{T}_{\varepsilon_0}}{f_{W_0}}(W_0) t_n \partial K \cdot \frac{w-W_0}{h} + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) = \\ &= \frac{1}{h^{d+1}} P \frac{\hat{T}_{\varepsilon_0}}{f_{W_0}}(W_0) t_n \partial K \cdot \frac{w-W_0}{h} + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \text{ uniformly over } W. \end{aligned} \quad (34)$$

Similar arguments as in Eq. 27 ensure that

$$\frac{1}{h^{d+1}} P \frac{\hat{T}_{\varepsilon_0}}{f_{W_0}}(W_0) t_n \partial K \cdot \frac{w-W_0}{h} = O_P \left( \frac{\hat{T}_{\varepsilon_0}}{f_{W_0}} t_n \right). \quad (35)$$

Using Eq. 35, 34, 33, 32, and 31, we obtain

$$\partial \hat{T}_{\varepsilon_1}(w) = O_P \left( \frac{n^{-1/2}}{h^{d+2} \tau_n} + \frac{\mathbb{P}_n |\hat{t}_n - t_n|}{h^{d+1} \tau_n} + \frac{\|(\hat{W} - W_0) t_n\|_\infty}{\tau_n} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0}) t_n\|_\infty}{\tau_n} \right) \text{ uniformly over } W. \quad (36)$$

and by proceeding similarly as for the previous display,

$$\hat{T}_{\varepsilon_1}(w) = O_P \left( \frac{n^{-1/2}}{h^{d+1} \tau_n} + \frac{\mathbb{P}_n |\hat{t}_n - t_n|}{h^d \tau_n} + \frac{\|(\hat{W} - W_0) t_n\|_\infty}{\tau_n} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0}) t_n\|_\infty}{\tau_n} \right) \text{ uniformly over } W. \quad (37)$$

To show the result for a general number of boosting iterations, say  $b$ , we will proceed by a simple induction. So, assume that for a generic  $b < B$  we have that uniformly over  $W$ ,

$$\begin{aligned} \partial \hat{T}_{\varepsilon_b}(w) &= O_P \left( \frac{n^{-1/2}}{h^{d+2} \tau_n^b} + \frac{\mathbb{P}_n |\hat{t}_n - t_n|}{h^{d+1} \tau_n^b} + \frac{\|(\hat{W} - W_0) t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0}) t_n\|_\infty}{\tau_n^b} \right), \\ \hat{T}_{\varepsilon_b}(w) &= O_P \left( \frac{n^{-1/2}}{h^{d+1} \tau_n^b} + \frac{\mathbb{P}_n |\hat{t}_n - t_n|}{h^d \tau_n^b} + \frac{\|(\hat{W} - W_0) t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0}) t_n\|_\infty}{\tau_n^b} \right), \text{ and} \\ \|(m_{\hat{W}}^{[b]} - m_{W_0}) t_n\|_\infty &= O_P \left( \frac{n^{-1/2}}{h^{d+1} \tau_n^b} + \frac{\mathbb{P}_n |\hat{t}_n - t_n|}{h^d \tau_n^b} + \frac{\|(\hat{W} - W_0) t_n\|_\infty}{\tau_n^b} + \frac{\|(m_{\hat{W}}^{[0]} - m_{W_0}) t_n\|_\infty}{\tau_n^b} \right). \end{aligned}$$

The, by proceeding as in Eq. 32, 36, and 37, we can show that the property will also hold for  $b+1$ . By Eq. 32, 36, and 37 the Induction Assumption holds for  $b=1$ , so we conclude.  $\square$

**Lemma 25.** *Let Assumptions 1-10 hold. Then,*

- (i)  $\frac{1}{\sqrt{n}} \mathbb{P}_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \partial K \cdot \frac{w-\hat{W}^*}{h} = O_P(1)$  uniformly over  $W$ ;
- (ii) If  $\hat{g}_n$  is a sequence of functions defined on  $\mathbf{X}$ ,  $\mathbf{F}_n$  is a sequence of classes of uniformly bounded functions defined on  $\mathbf{X}$ ,  $\log N_{[]}(\delta, \mathbf{F}_n, \|\cdot\|_\infty) \leq C\delta^{-v}$  with  $v \in (0, 2)$ ,  $P^*(\hat{g} \in \mathbf{F}_n) \xrightarrow{P} 1$ , and  $\|\hat{g}_n t_n\|_\infty = o_P(1)$ , then  $\frac{1}{\sqrt{n}} \mathbb{P}_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \hat{g}_n \psi = o_P(1)$  uniformly over  $\mathbf{T}$ , where  $\psi$  satisfies condition CL.
- (iii) If  $\hat{g}_n^*$  is a sequence of function defined on  $\mathbf{W}$  such that  $P(\hat{g}_n^* \in \mathbf{C}(\mathbf{W})) \rightarrow 1$  and  $\|\hat{g}_n^* t_n\|_\infty = o_P(1)$ , then  $\frac{1}{\sqrt{n}} \mathbb{G}_n \hat{g}_n^*(W_0) t_n \partial K \cdot \frac{w-\hat{W}^*}{h} = o_P(1)$  uniformly over  $W$ , where  $\mathbf{G}_n := \frac{1}{\sqrt{n}}(\mathbb{P}_n - P)$ .

*Proof.* For Point (i), define  $\tilde{Y}^* := 1' W_{0,1}(X) \geq u_1^* \cdot 1' \frac{m_{W_0}(W_0)}{W_{0,1}}(X) \geq u_2^*$ . By definition of  $Y^*$ ,

$$\begin{aligned} & \dots (Y^* - \tilde{Y}^*) t_n \dots_{L_2(P)} \leq \\ & \dots (W_1 - W_{0,1}) t_n \dots_{L_2(P)} + \dots \frac{\hat{m}_{\hat{W}}^{[B]}(\hat{W})}{\hat{W}_1} - \frac{m_{W_0}(W_0)}{W_{0,1}} \sum_{n \dots_{L_2(P)}} t_n \dots = o_P(1) \end{aligned}$$

where the last inequality follows from the convergence rates in Lemma 24, Assumption 6, and Assumption 8. Using the above display, the rates in Lemma 24, and Assumption 9, uniformly over  $W$ ,

$$\dots (Y^* - \hat{m}_{\hat{W}}^{[B]}(W)) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} \sum_h - (\tilde{Y}^* - m_{W_0}(W_0)) t_n \partial K \cdot \frac{w - W_0}{h} \sum_h = o_P(1).$$

From this point onward, the proof proceeds along the same arguments as Lemma 23, with Assumption 6-7 and Lemma 19 replaced by Assumption 9-10, Lemma 18, and Lemma 20.

For Point (ii), the assumption on the function  $\hat{g}$  and Lemma 24 ensure that

$$\dots (Y^* - \hat{m}_{\hat{W}}^{[B]}(W)) t_n \hat{g} \psi_t \dots_{L_2(P)} = o_P(1) \text{ uniformly over } T.$$

By  $P(\hat{g} \in C(W))_{n \rightarrow \infty} \rightarrow 1$ , Assumption 10, Lemma 18, and the previous display we can apply Lemma 20 to obtain the result of Point (ii).

The proof of Point (iii) proceeds along the same lines as the proof of Point (ii) in Lemma 19, using Assumption 9 and Lemma 20. □

**Lemma 26.** Under Assumptions 1-10,

$$\begin{aligned} (i) \dots \hat{T}_{\hat{\varepsilon}}^* \dots_{\infty, W} &= O_P \left( \frac{P_n \|\hat{t}_n - t_n\|}{h^d \tau_n^b} + \frac{n^{-1/2}}{h^{d+1} \tau_n^b} + \frac{\|(\hat{m}_{\hat{W}}^{*[0]} - \hat{m}_{\hat{W}}^{[0]})\|_{n, \infty}}{\tau_n^b} + \frac{\|(\hat{W}^* - \hat{W})\|_{n, \infty}}{\tau_n^b} + \frac{1}{\tau_n^b} \frac{\hat{T}_{\varepsilon_B} t_n}{f_{W_0}} \sum \right); \\ (ii) \dots \partial \hat{T}_{\hat{\varepsilon}_b}^* \dots_{\infty, W} &= O_P \left( \frac{P_n \|\hat{t}_n - t_n\|}{h^{d+1} \tau_n^b} + \frac{n^{-1/2}}{h^{d+2} \tau_n^b} + \frac{\|(\hat{m}_{\hat{W}}^{*[0]} - \hat{m}_{\hat{W}}^{[0]})\|_{n, \infty}}{\tau_n^b} + \frac{\|(\hat{W}^* - \hat{W})\|_{n, \infty}}{\tau_n^b} + \frac{1}{\tau_n^b} \frac{\hat{T}_{\varepsilon_B} t_n}{f_{W_0}} \sum \right); \\ (iii) \|(\hat{m}_{\hat{W}^*}^{*[1]} - \hat{m}_{\hat{W}}^{[0]}) t_n\|_{\infty} &= O_P \left( \frac{P_n \|\hat{t}_n - t_n\|}{h^d \tau_n^b} + \frac{n^{-1/2}}{h^{d+1} \tau_n^b} + \frac{\|(\hat{m}_{\hat{W}}^{*[0]} - \hat{m}_{\hat{W}}^{[0]})\|_{n, \infty}}{\tau_n^b} + \frac{\|(\hat{W}^* - \hat{W})\|_{n, \infty}}{\tau_n^b} + \frac{1}{\tau_n^b} \frac{\hat{T}_{\varepsilon_B} t_n}{f_{W_0}} \sum \right) \end{aligned}$$

*Proof.* By definition of  $\hat{T}_{\hat{\varepsilon}_b}^*$

$$\begin{aligned} \partial \hat{T}_{\hat{\varepsilon}_b}^*(w) &= \frac{1}{n h^{d+1}} \sum_{i=1}^n \hat{\varepsilon}_{0,i}^* (\hat{t}_{n,i} - t_{n,i}) \partial K \cdot \frac{w - \hat{W}^*}{h} \sum + \frac{1}{h^{d+1}} P_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} \sum \\ &\quad - \frac{1}{h^{d+1}} P_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}}^{*[0]}(\hat{W}^*)) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} \sum = O_P \left( \frac{P_n \|\hat{t}_n - t_n\|}{h^{d+1}} + \dots \right) \\ &\quad + \frac{1}{h^{d+1}} P_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} \sum + \frac{1}{h^{d+1}} P_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}}^{*[0]}(\hat{W}^*)) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} \sum \\ &\quad \text{uniformly over } W, \end{aligned} \tag{38}$$

where in the second equality we have used the consistency in Assumption 9-10. By Lemma 25,

$$\frac{1}{h^{d+1}} \mathbb{P}_n (Y^* - \hat{m}_{\hat{W}}^{[B]}(\hat{W})) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} = O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \text{ unif. over } W. \quad (39)$$

Using the rates in Assumption 6, 7, 9, and 10, together with the Lipschitz continuity of  $\partial K$ ,

$$\begin{aligned} & \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{m}_{\hat{W}}^{[B]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[0]}(\hat{W}^*)) t_n \partial K \cdot \frac{w - \hat{W}^*}{h} = \\ & \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{m}_{\hat{W}}^{[0]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[0]}(\hat{W}^*)) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + \\ & \sum_{s=0}^{B-1} \mathbb{P}_n \frac{\hat{T}_{\hat{W}}^{\varepsilon_S}}{f_{\hat{W}}}(\hat{W}) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \text{ unif. over } W. \end{aligned} \quad (40)$$

Using the rates in Assumption 6, 7, 9, 10, together with Lemma 25

$$\begin{aligned} & \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{m}_{\hat{W}}^{[0]}(\hat{W}) - \hat{m}_{\hat{W}^*}^{*[0]}(\hat{W}^*)) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} = \\ & \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{m}_{\hat{W}}^{[0]}(W_0) - \hat{m}_{\hat{W}^*}^{*[0]}(W_0)) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + \\ & - \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{W}^* - \hat{W}) \partial m_{W_0}(W_0) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) = \\ & \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{m}_{\hat{W}}^{[0]}(W_0) - \hat{m}_{\hat{W}^*}^{*[0]}(W_0)) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + \\ & - \frac{1}{h^{d+1}} \mathbb{P}_n (\hat{W}^* - \hat{W}) \partial m_{W_0}(W_0) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} + O_P \left( \frac{n^{-1/2}}{h^{d+2}} \right) \text{ unif. over } W. \end{aligned} \quad (41)$$

Proceeding in the same way as for Eq. 27 ensures that the two leading terms on the RHS of the previous display are, respectively,  $O_P \left( \|\hat{m}_{\hat{W}^*}^{*[0]} - \hat{m}_{\hat{W}}^{[0]}\|_{t_n} \right)$  and  $O_P \left( \|\hat{W}^* - \hat{W}\|_{t_n} \right)$  uniformly over  $W$ .

Proceeding along the same lines as in Eq. 34 and 35 yields

$$\sum_{s=0}^{B-1} \mathbb{P}_n \frac{\hat{T}_{\hat{W}}^{\varepsilon_S}}{f_{\hat{W}}}(\hat{W}) t_n \partial K \cdot \frac{w - \hat{W}_0}{h} = O_P \left( \frac{\hat{T}_{\hat{W}}^{\varepsilon_B} t_n}{f_{W_0}} + \frac{n^{-1/2}}{h^{d+2}} \right) \text{ unif. over } W. \quad (42)$$

Plugging Eq. 39, 40, and 42 into 38 yields

$$\begin{aligned} \partial \hat{T}_{\hat{W}}^* (w) &= O_P \left( \frac{\mathbb{P}_n |\hat{W} - \hat{W}^*|}{h^{d+1}} + \frac{n^{-1/2}}{h^{d+2}} + \|\hat{m}_{\hat{W}^*}^{*[0]} - \hat{m}_{\hat{W}}^{[0]}\|_{t_n} + \|\hat{W}^* - \hat{W}\|_{t_n} + \frac{\hat{T}_{\hat{W}}^{\varepsilon_B} t_n}{f_{W_0}} \right) \\ & \text{unif. over } W. \end{aligned} \quad (43)$$

By a similar reasoning,

$$\begin{aligned} \hat{T}_{\hat{W}}^*(w) &= O_P \left( \frac{\mathbb{P}_n |\hat{W} - \hat{W}^*|}{h^d} + \frac{n^{-1/2}}{h^{d+1}} + \|\hat{m}_{\hat{W}^*}^{*[0]} - \hat{m}_{\hat{W}}^{[0]}\|_{t_n} + \|\hat{W}^* - \hat{W}\|_{t_n} + \frac{\hat{T}_{\hat{W}}^{\varepsilon_B} t_n}{f_{W_0}} \right) \\ & \text{unif. over } W. \end{aligned} \quad (44)$$

Using the recursive structure of  $\hat{m}_{\hat{W}^*}^{*[1]}$  and the above convergence rate,

$$\begin{aligned}
& ||(\hat{m}_{\hat{W}^*}^{*[1]} - \hat{m}_{\hat{W}}^{[0]})t_n||_\infty = \because (\hat{m}_{\hat{W}^*}^{*[0]} - \hat{m}_{\hat{W}}^{[0]})t_n + \frac{\hat{T}_{\varepsilon_0}^*}{f_{\hat{W}^*}} t_n \dots_\infty = \\
& O_{P^*} \cdot \frac{P_n ||\hat{T}_{\varepsilon_0}^*||_n}{h^d \tau_n} + \frac{n^{-1/2}}{h^{d+1} \tau_n} + \frac{||(\hat{m}_{\hat{W}^*}^{*[0]} - \hat{m}_{\hat{W}}^{[0]})t_n||_\infty}{\tau_n} + \frac{||(\hat{W}^* - \hat{W})t_n||_\infty}{\tau_n} + \frac{1}{\tau_n} \dots \frac{\hat{T}_{\varepsilon_B}^*}{f_{W_0}} t_n \dots_\infty \\
& \text{unif. over } W \text{ .}
\end{aligned}$$

The convergence rates for  $\partial T_{\varepsilon_b}^*$ ,  $\hat{T}_{\varepsilon_b}^*$ , and  $\hat{m}_{\hat{W}^*}^{*[b]}$  are obtained by proceeding along the same lines as in the proof of Lemma 24 and Eq. 39, 40, 41, 42, and 38.  $\square$